

基于支持向量机的红外光谱天然气分析系统^{*}

杨开武¹ 白 鹏^{2,3} 李 彦³ 金 伟⁴ 刘君华²

(1.西北工业大学自动化学院 2.西安交通大学电气工程学院 3.空军工程大学理学院 4.上海化工研究院)

杨开武等.基于支持向量机的红外光谱天然气分析系统.天然气工业,2007,27(11):113-115.

摘 要 针对天然气组分浓度分析过程中海量训练样本无法实现、天然气组分气体的主次特征吸收谱线重叠严重、气相色谱分析法分析速度慢、无法进行实时在线分析等问题,将支持向量机(SVM)这一新的信息处理方法与红外光谱分析法结合,应用于天然气组分浓度分析中,利用 SVM 处理小样本的能力解决海量样本问题;通过建立反映天然气光谱数据样本与天然气组分浓度关系的 SVM 校正模型,解决天然气组分气体主、次特征吸收谱线重叠严重的问题;利用红外光谱分析速度快的优点,缩短分析时间。设计研制了基于 SVM 和红外光谱的天然气分析系统。该系统使用傅立叶红外光谱仪获取天然气红外光谱数据样本,对红外光谱数据样本进行数据预处理后,通过 SVM 校正模型进行计算分析,得出天然气组分浓度。实验结果表明,该方法的最大偏差为 3.95%,与气相色谱分析法相比,具有分析速度快、可实时在线分析等优点。

主题词 支持向量机 红外光谱 天然气组分 浓度 校正模型 定量分析

衡量天然气性能的主要指标之一是甲烷、乙烷、丙烷、正丁烷、异丁烷等气体的组分浓度。天然气组分浓度在生产和测量过程中,存在如下的问题。

(1)天然气具有组分多、组分浓度范围大的特点。例如,如果每种组分气体浓度按 1% 的间隔标定 100 个点,则 7 组分天然气需 100⁷ 个天然气样本。实际应用中无法实现海量的天然气样本。

(2)甲烷、乙烷、丙烷、异丁烷、正丁烷、异戊烷、正戊烷等气体的主征吸收谱线为 3017、2965、2968、2967、2967、2966、2965 cm⁻¹;次征吸收谱线为 1305、1470、1472、1477、1464、1472、1464 cm⁻¹。可以看出,其主和次特征吸收谱线重叠严重^[1]。

(3)天然气组分气体浓度常用的分析方法为气相色谱分析法^[2]。由于该法采用间歇取样,分析速度慢,无法进行实时在线分析。

因此,急待研究更快速的天然气组分浓度分析方法。笔者将支持向量机^[2-4](Support Vector Machine, SVM)这一新的信息处理方法与红外光谱分析法^[5]结合,应用于天然气组分浓度分析中,设计并实现了基于 SVM 和红外光谱的天然气分析系统。

一、基本原理

基本原理是:利用 SVM 处理小样本的能力解决海量样本问题;通过建立反映天然气光谱数据样本与天然气组分浓度关系的 SVM 校正模型,解决天然气组分气体的主、次特征吸收谱线重叠严重的问题;利用红外光谱分析速度快的优点,缩短分析时间。

1. SVM 校正模型

分析方法的核心是 SVM 校正模型,其结构见图 1。图 1 中,输入的 $x_i = (x_1, x_2, \cdots, x_l)$ 为在扫描波长范围内 l 个光谱数据; $y_i = (y_1, y_2, \cdots, y_m)$ 为对应的天然气 m 个组分浓度; $K(x_i, x_j)$ 为核函数; b 为阈

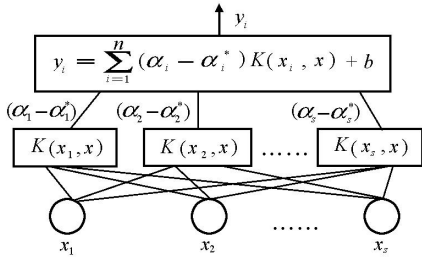


图 1 SVM 校正模型结构图

^{*} 本文为国家自然科学基金资助项目“基于多波长窄带光谱扫描技术的全光纤混合气体定量检测系统的研究”(编号:60772016)和陕西省科技计划资助项目“基于 SVM 的混合气体红外光谱多参量数字分析关键技术研究及应用”(编号:2007K05-05)研究成果。

作者简介: 杨开武,1963 年生,高级工程师,博士研究生;现从事天然气测量及可靠性研究工作。地址:(710072)陕西省西安市友谊西路 127 号。电话:(029)88389926,13709261150。E-mail: yangkaiwu@ qhgas. com

值; α 、 α^* 为 Lagrange 因子。

SVM 校正模型训练和检验流程如图 2 所示。

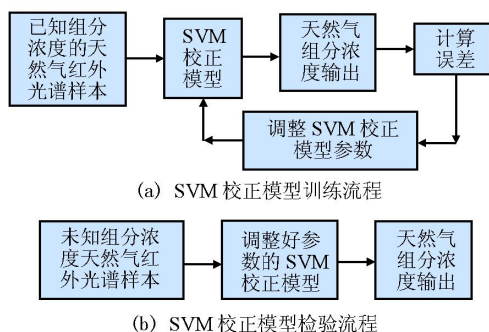


图2 SVM校正模型训练和检验流程图

先确定核函数、惩罚因子(C)及损失函数(ϵ)等参数,然后用一部分光谱数据样本对SVM校正模型进行训练,确定 α 和 b 。当输出组分浓度与期望值误差满足要求时,训练结束。用另一部分样本对SVM校正模型进行检验,如果满足误差要求,SVM校正模型最终确定。

2. 处理流程

(1) 天然气红外光谱数据样本的建立

1) 用已知浓度的标准气体配制天然气样本。天然气组分气体最小浓度为0.0%,组分气体最大浓度为100.0%,最小浓度间隔为0.0025%。

2) 扫描天然气样本。用红外光谱仪对已知组分浓度的天然气样本进行扫描,建立已知组分浓度的天然气光谱数据样本。实验所用红外光谱仪为Bruker公司的TENSOR 27型傅立叶变换红外光谱仪,扫描范围为4000~400 cm^{-1} ,扫描间隔为12 nm,扫描后得到5500个天然气红外光谱数据样本。用于训练和检验SVM校正模型。

光谱数据样本结构为: $\{(x_i, y_i), i=1, \dots, n\}$,其中 $x_i \in R^d$ 为第 i 个天然气红外光谱数据样本。 x_i 、 y_i 同前。

(2) 光谱数据样本预处理

对获得的天然气光谱数据样本进行数据归一扩展预处理,消除噪声和漂移的影响,提高天然气组分浓度分析的准确性。归一即统一标准;扩展即将归一的光谱数据进行放大,便于SVM校正模型对光谱数据微小变化的处理。可用下式来描述。

$$x_i = \frac{x_i}{x_{av}} \times A \quad x_{av} = \frac{1}{l} \sum_{i=1}^l x_i$$

式中: x_{av} 为天然气光谱数据样本的平均值, l 为光谱数据点的个数, A 为放大倍数。

(3) SVM校正模型的训练与检验

SVM校正模型训练和检验流程,如图2所示。

(4) SVM校正模型的实现

对于本文的天然气分析系统,采用图3所示的多层次多SVM校正模型结构实现SVM校正模型。

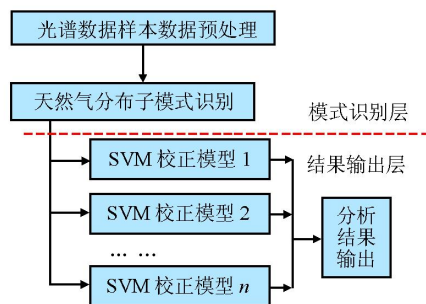


图3 多层次多SVM校正模型结构图

图3所示的结构中,模式识别层对输入的的光谱数据样本进行天然气分布子模式识别;结果输出层进行详细的计算。根据光谱数据样本所属的天然气分布子模式,分配给相应的SVM校正模型,计算出相应的结果。多层次多SVM校正模型结构,对复杂任务实现了分而治之的目的。

二、系统设计与实现

1. 总体结构

系统总体框架由傅立叶变换红外光谱仪、计算机、软件等部分组成,如图4所示。工作流程为:计算

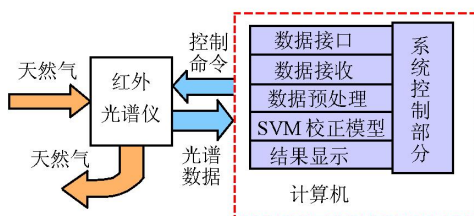


图4 系统框架图

机发出控制命令,红外光谱仪获得天然气红外光谱数据样本;计算机通过数据接口部分对光谱数据样本进行接收,光谱数据样本经数据预处理后由SVM校正模型进行计算分析,最后得出分析结果。上述流程在系统控制部分的控制下进行。

2. 系统集成实现

系统集成实现的硬件如图5所示。

红外光谱仪获取天然气光谱数据样本,通过SVM校正模型进行分析,得出天然气组分浓度。

软件部分采用模块化设计思想,其组成见图6。

红外光谱数据接收部分:根据控制命令,接收天然气样本红外光谱数据。数据库:将接收到的天然

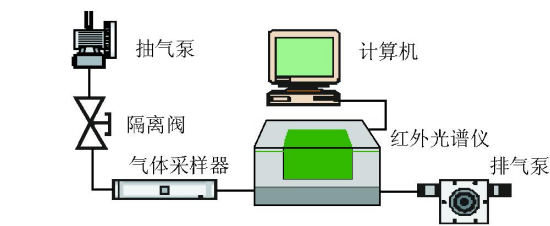


图 5 天然气分析系统硬件图

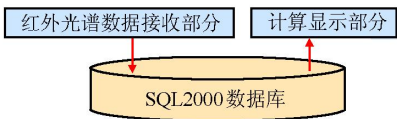


图 6 软件部分组成图

气样本红外光谱数据进行保存。计算显示部分:对数据库中保存的天然气样本红外光谱数据进行分析计算,计算结果通过显示器或打印机输出。

三、实验结果

选用线性核函数、惩罚因子(C)=8,进行归一扩展数据预处理,编号为奇数的光谱数据用于训练,编号为偶数的用于检验。

1. 实验数据

按照浓度范围,甲烷 95%~30%,乙烷 80%~10%,丙烷 60%~5%,异丁烷、正丁烷 40%~5%,异戊烷、正戊烷 30%~1%的模式,浓度最小间隔为 0.1%,采用标准气体配制混合气体样本 5500 个。将 5500 个光谱数据样本的一半用于训练 SVM 校正模型;另一半用于检验分析系统。将事先已知组分浓度的天然气作为真实值,结果见表 1。

表 1 用标准天然气测试的评估指标表 %

气体	最大绝对 误差值	最小绝对 误差值	平均绝对 误差值
甲烷	3.95	0.0	0.73
乙烷	3.83	0.0	0.42
丙烷	0.6	0.0	0.10
异丁烷	1.1	0.0	0.05
正丁烷	2.1	0.0	0.05
异戊烷	1.7	0.0	0.04
正戊烷	2.02	0.0	0.04

从上述实验可得出结论:利用天然气红外光谱数据样本,通过 SVM 校正模型进行天然气组分浓度分析方法的是可行的。

2. 实际应用

为了检验本分析系统,将一台经过计量标定的气相色谱分析仪作为标准计量仪器进行对比,将气相色谱仪的输出结果作为真实值,将分析系统的输出作为计算分析值,系统的分析结果如表 2 所示。

表 2 用标准仪器对比测试的评估指标表 %

气体	最大绝对 误差值	最小绝对 误差值	平均绝对 误差值
甲烷	3.65	0.01	0.14
乙烷	0.40	0.0	0.01
丙烷	0.88	0.0	0.04
异丁烷	0.06	0.0	0.01
正丁烷	0.11	0.0	0.01
异戊烷	0.17	0.0	0.01
正戊烷	0.20	0.0	0.02

现场应用时的分析速度小于 12 s,与气相色谱分析法相比较,分析速度有明显提高。

四、结 论

与气相色谱仪器相比,本文的分析系统具有如下功能:①具有在线联机实时分析功能,适合于现场应用;②能进行离线分析,适合于实验室分析;③光谱数据样本可保存到数据库中,为历史数据分析提供方便;④通过光谱图、报表等方式显示、保存天然气组分浓度分析结果。

从现场应用来看,分析系统的 SVM 校正模型使用良好,分析精度满足期望的要求,为天然气组分浓度分析提供了新的技术手段,为支持向量机的应用开辟了新的领域,丰富了仪器的信息处理方法,具有推广应用前景。

参 考 文 献

[1] 徐文渊,蒋长安.天然气利用手册[M].北京:中国石化出版社,2002.

[2] VAPNIK V N. The nature of statistical learning [M]. Berlin:Springer,1995.

[3] 张学工.关于统计学习理论与支撑向量机[J].自动化学报,2000,26(1):32-42.

[4] 姚凯丰,陆文凯,丁文龙,等.一种基于 SVM 特征选择的油气预测方法[J].天然气工业,2004,24(7):36-38.

[5] 陆婉珍,袁洪福,徐广通,等.现代近红外光谱分析技术[M].北京:中国石化出版社,2000.

(修改回稿日期 2007-09-25 编辑 赵 勤)