

基于不同波段对成品汽油的模式识别分析

姜黎^{a,b} 张军^{a,b} 陈哲^{a,b} 余谦^{a,b} 王京华^c

^a[广东省高等学校光电信息与传感技术重点实验室(暨南大学) 广州市 510632]

^b(暨南大学理工学院光电工程系 广州市黄埔大道西 601 号 510632)

^c(中国石化总公司广州分公司 广州市 510726)

摘 要 为了实现对输油管道上顺序输送不同牌号汽油(90#, 93#, 97#)时汽油间界面的区分, 根据近红外光谱分析的基本原理, 采用主成分分析-马氏距离(PCA-MD)模式识别方法, 分析了700—1100nm 波段和1100—1700nm 波段3种不同牌号成品汽油的近红外光谱, 并根据分析结果对汽油进行了分类。结果表明, 使用1100—1700nm 波段的汽油光谱分类结果较好, 该方法可进一步鉴别汽油的质量。

关键词 输油管道界面检测; 主成分分析-马氏距离; 模式识别; 成品汽油

中图分类号: O 657. 33; O 433. 1 文献标识码: A 文章编号: 1004-8138(2010)03-1208-05

1 引言

成品油顺序输送的界面切割一直是输油管道检测中的突出问题^[1], 传统的检测方法是进行密度^[2]或者折射率检测^[3]来确定混油界面。这些方法虽能够达到界面切割的目的, 但存在着对于2种汽油的界面定位精度差、检测效率低等不足。根据已有经验, 石油产品的性质、组成分析以及在线分析已广泛采用近红外光谱分析的方法。为此, 本文提出采用近红外光谱分析的方法对汽油进行研究。近红外光谱分析是一种间接的分析技术, 通过建立校正模型, 实现对未知样本的定性或定量分析^[4]。近红外(NIR)光谱分析技术的最大特点是分析周期短、成本低、分析结果的重现性好。国内外对近红外光谱在汽油的研究中做了大量的工作^[5], 如田高友^[6]等人研究了不同牌号汽油的近红外光谱识别技术, 研究结果表明, 利用近红外光谱技术对90#和93#汽油烃族组成的差异进行识别是可行的。史月华^[7]等人, 将主成分回归残差神经网络校正算法用于近红外测定汽油辛烷值的预测模型校正。张其可^[8]等人提出基于近红外光谱的汽油牌号快速识别算法, 主要包括预处理、特征提取和分类建模几部分, 并且比较了各种分类方法的识别能力。Kim M^[9]基于近红外光谱, 采用主成分和贝叶斯统计方法对石油产品(汽油、柴油等)进行了实时分类。付大友^[10]等用聚类分析方法对汽油质量进行了研究。尽管如此对于输油管线油品界面分析方面, 光谱分析的应用研究报道尚不多见。

本文针对输油管道检测中较为突出的问题——不同牌号汽油顺序输送时界面难于区分, 基于近红外光谱分析技术, 将主成分分析-马氏距离(PCA-MD)模式识别方法应用于700—1100nm 波

国家 863 项目(2009AA 04Z315)

联系人, 手机: (0)13265044828; E-mail: zhangz-line@sina.com

作者简介: 姜黎(1983—), 女, 湖南省邵东县人, 硕士研究生, 主要从事光谱分析及精密仪器制造方面的研究工作。

张军(1968—), 女, 长春市人, 副研究员, 博士, 主要从事工程光学及精密仪器制造方面的研究工作。

收稿日期: 2010-01-06; 接受日期: 2010-01-31

段和1100—1700nm 波段汽油的快速分析。首先通过PCA 提取成品汽油近红外光谱的主成分信息, 并结合判别式聚类分析分类方法建立各类汽油样本的分类模型, 利用这些模型实现对未知汽油样本的快速识别。

2 方法与原理

近红外光谱识别不同牌号汽油主要包括: 光谱测量、光谱预处理、特征提取与分类等步骤, 具体结构如图1 所示。

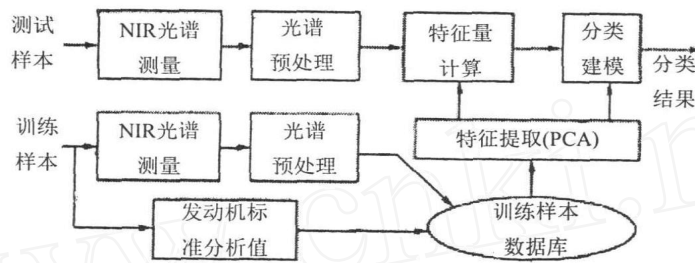


图 1 近红外光谱分析技术自动识别过程^[8]

主成分分析-马氏距离(PCA-MD)法是美国材料与试验协会(ASTM)制定的近红外光谱模式识别的标准方法。该方法基本思想是*i*类的校正集光谱经过PCA分析, 选取分类能力强的主成分得分作为识别特征, 再计算出*i*类各样品到*i*类中心的马氏距离, 并确定*i*类样品的马氏距离范围MD threshold。对于未知样品, 首先计算该样品到*i*类中心的马氏距离MDunkow n, 如果MDunkow n小于MD threshold, 则该样品属于*i*类, 反之不属于*i*类。通常, MD threshold = MD average + $k \times MDSTD$ 。其中MD average为*i*类校正集样品的马氏距离平均值, MDSTD为*i*类样品的马氏距离平均偏差, k 为阈值参数。阈值过小, 存在*i*类样品漏识别可能性增加, *i*类样品错误识别可能性减少; 反之, 阈值过大, 则*i*类样品漏识别可能性减少, *i*类样品错误识别可能性增加。因此*k*值的选择需要综合考虑, 通过调节*k*值可以确定最佳阈值。由于PCA-MD方法易受光谱预处理和变量选择的影响, 本文PCA-MD处理过程依次为: (1) 原始光谱进行标准归一化处理, (2) 进行PCA(最大主因子数为10)分析, (3) 评价10个主成分对3种不同牌号汽油的识别能力的影响, (4) 选择识别能力最强的前3个主成分作为识别特征变量。

3 实验部分

3.1 实验平台

本实验采用CCD固定光路阵列探测器型近红外光谱分析仪(北京英贤仪器实业有限公司制造)^[11], 测定各汽油的NIR光谱。光谱仪的探测器分为Si探测器和InGaAs探测器2种, 有效波长范围为700—1100nm和1100—1700nm。相关的数据分析软件均用Matlab7.0编写。

3.2 样本集

样本数据由中石化公司提供, Si探测器光谱仪采集的数据包括不同牌号的655个汽油样本, InGaAs探测器光谱仪采集的数据包括不同牌号的338个汽油样本。样本集包括各汽油样本的原始NIR光谱数据与采用ASTM-CFR标准方法研究测定的辛烷值指标。样本集分为3个类别: 90#、93#、97#汽油。为保证分类的公平合理, 各牌号的训练集均由随机选取的30个标准样本组成, 剩下的样本作为测试集用于评估方法的有效性(如表1所示)。

表 1 样本数据集划分

样品类别	90# (个)	93# (个)	97# (个)	总计(个)
700—1100nm	训练集	30	30	30
	测试集	249	171	145
1100—1700nm	训练集	30	30	30
	测试集	64	122	62

4 结果与分析

4.1 数据预处理

由于不同样本测量时,光源强度差别以及仪器背景噪声或漂移等因素对检测信号的影响,通常需要对测量的光谱数据进行预处理。光谱数据预处理方法包括矢量归一化(SNV)以及多项式卷积平滑等多种方法。通过对原始光谱进行一阶求导、二阶求导、多元散射校正、SNV及多项式卷积平滑等多种预处理方法比较,得出光谱采用SNV以及一阶导数处理后,回归模型的相关系数最大,均方残差最小。故本文对光谱数据均采用SNV以及一阶导数预处理方法。

4.2 光谱特征提取

在建模之前,首先对参与建模的光谱信息进行选择是一种有效的优化模型方法,对减少噪音信号的影响、提高运算速率和模型的稳定性是有益的。又由于采用全谱计算时,计算量大而且过程繁冗,有些光谱区域中样品光谱信息较弱,光谱与样品组成或性质间缺乏相关关系。因此,在光谱建模过程中,需对波长点进行选择,删除那些对建模贡献小或有干扰作用的波长点,以达到优化模型的目的。本文采取PCA法对预处理后的光谱数据进行数据压缩和信息提取。图2为采用PCA法对不同牌号汽油训练集吸收光谱建模后的前3个主成分两两组合的得分图。

从图2中可以看出,PC1和PC2组合所包含的分类信息最丰富,类别分隔基本没有重叠区,可以有效分类出不同牌号的汽油。

4.3 700—1100nm 波段和 1100—1700nm 波段成品汽油分类结果

采用PCA-MD模式识别方法对预处理和特征提取后的训练集样品光谱,建立3种牌号(90#、93#和97#)汽油的识别校正模型。校正模型建立后,采用分类错误率作为判别模型好坏的指标。分类错误率等于测试集一类样品判别错误的样品数占该类测试样品总数的百分比。分类错误率越低代表分类效果越好。分别计算不同牌号汽油的测试样本在700—1100nm波段和1100—1700nm波段的分类错误率(如表2所示)。

表 2 700—1100nm 和 1100—1700nm 波段对汽油牌号的分类预测结果

分类错误率(%)	90#	93#	97#
700—1100nm 波段	17.20	34.32	12.0
1100—1700nm 波段	12.5	13.93	12.9

根据表2的结果,对比2个波段的分类错误率,可以看出使用1100—1700nm波段光谱进行成品汽油分类的结果,优于使用700—1100nm波段光谱的分类结果。为了进一步验证表2的结果,另分别取37个90#、28个93#、33个97#成品汽油,在700—1100nm和1100—1700nm波段采用同样的条件进行光谱测量、光谱预处理,仍然采用上面的识别模型进行判别。此时得出的预测结果如表3。

表 3 新测试集汽油牌号分类结果

分类错误率(%)	90#	93#	97#
700—1100nm 波段	0	39.28	3.03
1100—1700nm 波段	0	14.28	3.03

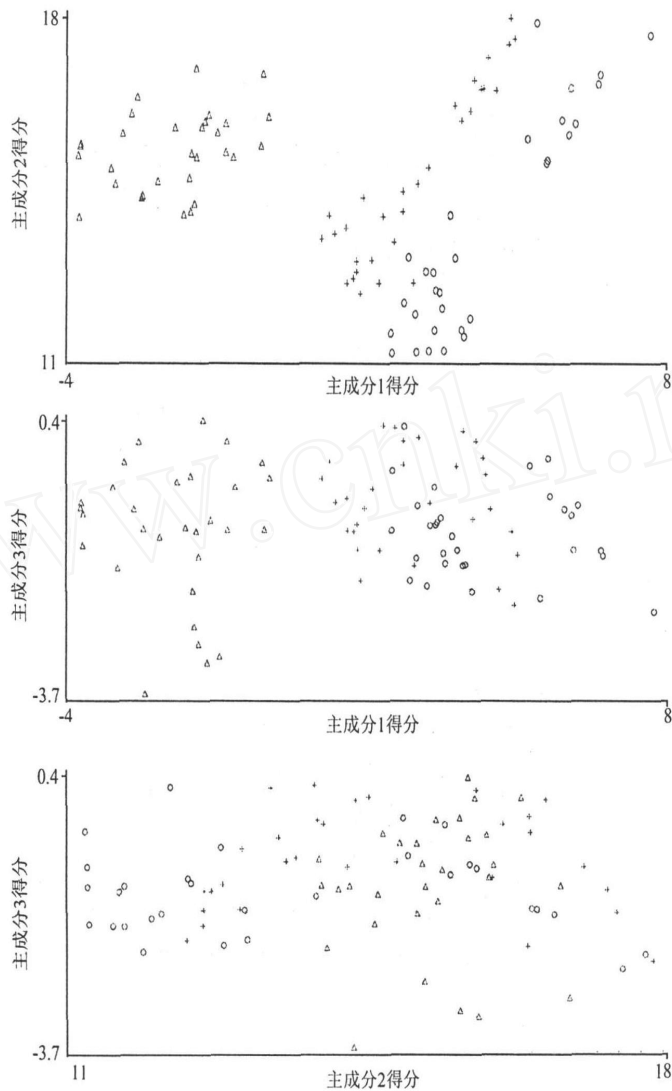


图 2 采用 PCA 法后训练集样本前 3 个主成分两两得分组合图

表 3 结果表明用同一批成品汽油光谱数据, 在不同波段下进行模式识别分析, 1100—1700nm 波段的模式识别能力优于 700—1100nm 波段, 从而有力的验证了表 2 的结论。由于在管道中传输的汽油混合界面是两两相互混合的, 因此选择在 1100—1700nm 波段采用 PCA-MD 方法分别两两建立识别模型, 即 93# 和 97#, 90# 和 93#, 90# 和 97# 组成的识别模型。选表 1 的样品组成训练集和测试集。得到的结果如表 4 所示。表 4 说明当两两建模时, 90# 和 93# 汽油容易混淆, 较难识别。93# 和 97# 汽油或者是 90# 和 97# 汽油较容易区分。

表 4 1100—1700nm 波段两两建模的汽油牌号分类结果

分类错误率(%)	90# 和 93# 建模结果		93# 和 97# 建模结果		90# 和 97# 建模结果	
	90#	93#	93#	97#	90#	97#
1100—1700nm 波段	10.9	7.4	0	3	0	2.4

5 结论

本文针对不同牌号汽油顺序输送时界面难于区分的问题,建立了基于NIR 光谱的汽油牌号快速识别算法,算法主要包括数据预处理、特征提取和分类建模3部分。文中主要采用PCA-MD 模式识别方法在700—1100nm 和1100—1700nm 光谱范围内分别建立成品汽油的定性检测模型,并对这两种模型的识别能力进行了比较。实验结果表明:在1100—1700nm 波段进行模式识别的性能优于700—1100nm 波段。为了降低汽油分类错误率,实现对管道中传输汽油混油界面的检测,建立通用的校正模型,进一步的工作正在进行中。

参考文献

- [1] 钟仕荣,扬金剑,王建华.原油顺序输送的混油界面跟踪与切割[J].油气储运,2005,24(8):7—9
- [2] 田西宁,王卫强,常青等.一种光学界面检测仪在成品油管道上的应用[J].管道技术与设备,2008,14(6):20—22
- [3] 诸葛晶晶,曾周末,陆丽等.基于光学传感器的成品油识别方法[J].光学精密工程,2009,17(6):1479—1484
- [4] 陆婉珍.现代近红外光谱分析技术[M].北京:中国石化出版社,2007.
- [5] 严衍祿,赵龙莲,韩东海等.近红外光谱分析基础与应用[M].北京:中国轻工业出版社,2005
- [6] 田高友,熊春华,刘慧颖.近红外光谱技术识别车用汽油的应用研究[J].现代科学仪器,2006,15(5):84—87.
- [7] 史月华,陆勇,徐光明等.主成分回归残差神经网络校正算法用于近红外光谱快速测定汽油辛烷值[J].分析化学,2001,29(1):87—91.
- [8] 张其可.基于近红外光谱的汽油牌号识别与辛烷值测定[D].杭州:浙江大学,2006
- [9] Kim M, Lee Y, Han C. Real-Time Classification of Petroleum Products Using Near-Infrared Spectra[J]. *Comp. Chem. Eng.*, 2000, 24(2): 513—517.
- [10] 付大友,袁东,谭文渊等.聚类分析方法用于汽油质量研究[J].吉林化工学院学报,2005,22(2):6—10
- [11] 袁洪福,褚小立,陆婉珍.一种新型CCD在线近红外光谱分析仪的研制[J].分析化学,2004,32(2):255—261.

Study on Pattern Recognition Analysis Method for Gasoline Based on Different Spectral Bands

J IANG L i^{a,b} Z HANG Jun^{a,b} C HEN Z he^{a,b} Y U Q ian^{a,b} W ANG J ing-Hua^c

a[Key Laboratory of Optoelectronic Information and Sensing Technologies of Guangdong Higher Educational Institutes, (Jinan University), Guangzhou 510632, P. R. China]

*b*Department of Optoelectronic Engineering, Jinan University, Guangzhou 510632, P. R. China)

c(China Petrochemical Corporation, Guangzhou Branch, Guangzhou 510726, P. R. China)

Abstract The near infrared spectroscopy of gasoline has been analyzed in order to research the interfaces of gasoline (90#, 93#, 97#) in pipeline. The principal component analysis-Mahalanobis Distance (PCA-MD) pattern recognition method is used according to the basic principles of near-infrared spectroscopy. The calibration models of three kinds of gasoline have been built on 700—1100nm and 1100—1700nm spectra respectively. The results show that the classification ability of model on 1100—1700nm range is better than the model on 700—1100nm spectrum. The calibration method can further identify the quality of gasoline.

Key words Pipeline Laying Interface Detection; PCA-MD; Pattern Recognition; Gasoline