

# 基于光谱技术鉴别机油品种的新方法

周子立<sup>1,3</sup> 蒋璐璐<sup>2</sup> 谈黎虹<sup>2</sup> 何 勇<sup>3</sup> 李晓丽<sup>3</sup> 邵咏妮<sup>3</sup>

(<sup>1</sup> 浙江机电职业技术学院, 浙江 杭州 310053; <sup>2</sup> 浙江经济职业技术学院, 浙江 杭州 310018)  
<sup>3</sup> 浙江大学生物工程与食品科学学院, 浙江 杭州 310029

**摘要** 提出了一种用可见-近红外透射光谱技术快速鉴别机油品种的新方法,应用可见-近红外光谱仪测定三种机油的光谱曲线,然后用主成分分析法对不同品种的机油样本进行聚类分析,并获取机油可见-近红外光谱的特征信息,再结合多类判别分析技术建立机油品种鉴别的模型,对经过预处理的光谱数据进行主成分分析。结果表明,以样本在第一主成分和第二主成分上的得分做出的二维散点图,对不同种类机油具有很好的聚类,能定性区分不同种类机油;经过主成分分析得到的前 8 个主成分的累积可信度已达 95.38%,说明这 8 个变量能够代表绝大部分原始光谱的信息。从 180 个样本中随机抽取 150 个样本用于建立多类判别分析品种鉴别模型,余下的 30 个样本用于验证。对未知的 30 个样本进行品种预测,准确率为 100%。证明本方法具有明显的分类和鉴别作用,为不同品种的机油鉴别提供了一种新方法。

**关键词** 光谱学;机油;近红外光谱;主成分分析;多类判别分析;鉴别

**中图分类号** O 657.3; TE626.3+2 **文献标识码** A **doi**: 10.3788/AOS20092908.2203

## Discrimination of Oil Varieties by Using Near Infrared Spectral Technology

Zhou Zili<sup>1,3</sup> Jiang Lulu<sup>2</sup> Tan Lihong<sup>2</sup> He Yong<sup>3</sup> Li Xiaoli<sup>3</sup> Shao Yongni<sup>3</sup>

<sup>1</sup> Zhejiang Institute of Mechanical & Electrical Engineering, Hangzhou, Zhejiang 310053, China

<sup>2</sup> Zhejiang Technology Institute of Economy, Hangzhou, Zhejiang 310018, China

<sup>3</sup> College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, Zhejiang 310029, China

**Abstract** A new method for discrimination of oil types by means of near infrared spectroscopy (NIRS) was developed. First, the characteristic spectrums of oil were got through principal component analysis (PCA). The result of the analysis suggests that the reliabilities of first 8 principal components are more than 95.38%. The 2-dimensional plot was drawn with first and second principal components, which indicates that it is a good clustering analysis for classification of oil varieties. Several variables compressed by PCA were used as inputs of multiple discriminant analysis (MDA). 150 samples from three varieties were selected randomly, then they were used to build discrimination model. 30 unknown samples were predicted by this model, and the recognition rate is 100%. This model is reliable and practicable. It could offer a new approach to the fast discrimination of oil types.

**Key words** spectroscopy; oil; near infrared spectra; principal component analysis (PCA); multiple discriminant analysis (MDA); discrimination

## 1 引 言

机油是保证发动机正常运转,减少机件间摩擦的必需润滑剂,其内在品质的检测及品种鉴别一直

是机油生产研究的重要课题。国内外学者已在机油品质对减少机械磨损方面做过大量研究<sup>[1~3]</sup>,也有学者在机油添加剂技术及测量机油粘度方面进行了

收稿日期: 2008-10-27; 收到修改稿日期: 2008-12-20

基金项目: 十一五国家科技支撑项目(2006BAD10A0902),浙江省重大科技攻关项目(2005C12029)、浙江省教育厅资助项目(20071275)和浙江机电职业技术学院孵化基金项目(04090331)资助课题。

作者简介: 周子立(1963—),男,高级工程师,主要从事计算机信息处理及计算机应用技术等方面研究。

E-mail: zhouzili88@163.com

研究<sup>[4,5]</sup>,但对机油的品种鉴别方面研究还很少。研究一种简单、快速的机油品种鉴别方法,在目前车用机油种类繁多,良莠不齐,亟待整治的市场背景下是很有必要的。利用这种简单、快速的鉴别方法即可防止假冒产品以次充好,也可减少甚至避免劣质机油进入市场。本文以近红外光谱技术为基础研究机油的品种鉴别,相对与红外光谱技术研究具有容易实现、快速、高效、低成本的特点,便于此项鉴别技术的普及与推广。

现代可见-近红外光谱分析技术,可充分利用全谱段或多波长下的光谱数据进行定性或定量分析。由于可见-近红外光谱技术分析具有速度快、效率高、成本低、测试重现性好、测量方便等特点,已经被越来越多地应用于食品工业、石油化工、农业等领域。有学者研究利用近红外光谱技术检测水果内部品质<sup>[6]</sup>,茶叶中茶素含量<sup>[7]</sup>,判别海面溢油种类<sup>[8]</sup>、制动液品牌<sup>[9]</sup>、车用汽油标号<sup>[10]</sup>、预测黄酒的糖度<sup>[11]</sup>等。但是可见-近红外波谱段具有信息量大、有噪声干扰、波谱重叠等特征,如果直接运用原始光谱数据进行判别分析往往导致模型精度低,模型稳定性差。应用主成分分析(PCA)结合多类判别分析方法挖掘光谱中的有用信息,在不丢失主要光谱信息的前提下选择为数较少的新变量来代替原来较多的变量,解决了由于谱带的重叠而无法分析的困难。本文采用主成分分析和多类判别分析法建立机油不同品种的近红外光谱鉴别模型,现实了机油品种的快速鉴别。

## 2 材料与方法

### 2.1 仪器设备

实验使用美国 ASD (Analytical Spectral Device) 公司的 Handheld FieldSpec 光谱仪,其光谱采样间隔(波段宽) 1.5 nm,测定范围 325 ~ 1075 nm,扫描次数 30 次,分辨率 3.5 nm,探头视场角为 90°。光源是与光谱仪配套的 14.5 V 卤素灯。光谱数据以 ASCLL 码形式导出进行处理,分析软件为 ASD View Spec Pro, Unscramble 和 DPS (data procession system for practical statistics)。

### 2.2 样本来源及光谱的获取

从市场购买三种机油,浙江壳牌化工石油有限公司生产的壳牌白喜力(Shell oil),中国石化公司润滑油分公司生产的长城福星(Great Wall oil),广西玉柴高级润滑油有限责任公司生产的玉柴牌机油(Yuchai oil)。为保证实验样本的均匀性,每种机油

选用了 10 个不同的生产批次和日期。机油用直径 65 mm 高度 14 mm 的透明器皿盛装。装满 1 个器皿作为一个实验样本。三种机油按不同生产批次和日期各取 60 个样本,共 180 个样本。全部样本随机分成建模集和预测集,建模集有 150 个样本(每个品种 50 个样本),预测集有 30 个样本(每个品种 10 个样本)。光谱仪预热 20 min 经白板校准后进行测试。如图 1 所示,光谱仪置于机油样本的下方,探头距离样本底部 10 mm,探头视场角为 90°,光线经过机油样本垂直透射在光谱仪探头上,光谱仪对每一个样本扫描 30 次,取其采样透射光谱平均值,并分别保存。

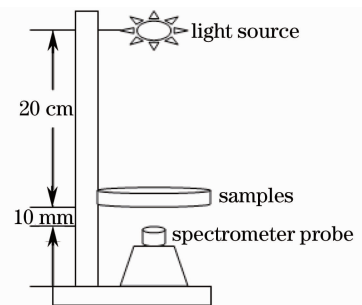


图 1 实验样本测试情况

Fig. 1 The test of oils in experiment

### 2.3 光谱数据预处理

为了去除来自高频随机噪音、基线漂移、样本不均匀、光散射等影响,需要进行光谱预处理来消除噪音。采用 Move average 平滑法,选用平滑点数为 9,此时能很好滤除各种因素产生的高频噪音,再进行二阶求导处理,消除基线漂移。光谱曲线在首端和末端有较大噪音,为了减少噪音,提高精度,去掉前 75 nm 和后 75 nm 波段,选用 400 ~ 1000 nm 光谱范围进行研究<sup>[12]</sup>。

### 2.4 多类判别分析法

多类判别分析法是一种集“有效特征选择与状态识别”功能于一身的统计分析法,判别分析是一种广泛应用的多元统计方法,主要是判别未知样本应划归哪一个已知总体。在判别未知样本前,首先要研究不同总体的性质和特征,根据已知总体的多种观测指标建立判别函数,并以它作为样本划归某一总体的依据。最终把对分类起作用的多个指标建立数学判别式,最后将待判别样品的有关数值代入判别函数,即能判别该样本的类别<sup>[13]</sup>。

## 3 实验结果与讨论

### 3.1 机油品种模型的建立

三种机油典型可见-近红外透射光谱曲线如图

2 所示。图中横坐标为波长范围 325~1075 nm,纵坐标为光谱漫反射率。从图 2 中可以看出,三种机油在 600~700 nm 光谱范围内有很大的差异,在 700~1075 nm 范围光谱曲线具有相同的趋势,但是存在较大的基线漂移,运用二阶求导消除基线漂移,使不同品种的谱线差异更明显。所以,不同品种机油的光谱图有明显区别,并具有一定的特征性和指纹性,这一差异为机油的不同品种鉴别奠定了数学基础<sup>[14]</sup>。不同品种机油的光谱曲线差异,不仅跟它们的视觉特征有关而且跟它们的内部品质譬如粘度等密切相关。

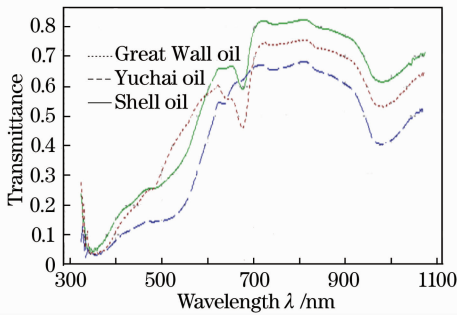


图 2 三种机油的典型可见-近红外光谱透射图  
Fig. 2 Visible-near infrared transmittance of three different oils

### 3.2 主成分分析对不同品种机油进行聚类

主成分分析不仅能够对降低数据维数<sup>[15]</sup>,将多波长下的光谱数据压缩到有限的几个因子空间内,使数目较少的新变量能最大限度的表征原变量的数据结构特征,并不丢失信息。而且能够通过样本在各因子空间的得分确定所属的类别,所以新变量能够更加形象的表征原样本的品质差异、品种区别等。光谱数据经预处理并选择光谱范围后,对其做主成分分析。以样本在第一主成分和第二主成分上的得分作图,结果见图 3。

图 3 为主成分 1、2 所作的二维散点得分图,图

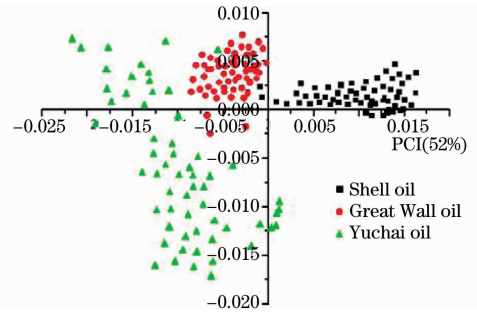


图 3 180 个样本的主成分 1 和主成分 2 得分图  
Fig. 3 Scores plot obtained from the PCA (PC1 and PC2) of 180 samples

中横坐标表示每个样本的第一主成分得分值,纵坐标表示每个样本的第二主成分得分值。图 3 中三种机油明显分成三类,说明主成分 1、2 对三种机油有较好的聚类作用。从图 3 中可以看出,长城机油 60 个样本聚合度很好,紧密地分布在图 3 中坐标系的第二象限附近;壳牌机油 60 个样本与其他两个品种的样本分界很清楚,它们都位于图 3 中坐标系的第二象限附近即坐标系中纵坐标的右边,而其它两个品种的样本大都位于坐标系中纵坐标的左边;玉柴机油的 60 个样本的聚合度没有前两个品种好,它们分布在图 3 坐标系中的第三、四象限,但是没有跟另两个品种混合起来,它们之间的分界线清楚。分析表明主成分分析对三种机油有一定的聚类作用,能定性区分不同品种的机油。

### 3.3 基于多类别判别分析建立品种鉴别模型

光谱波段从 400~1000 nm 共有 600 个点,但是采用 600 个点计算时,计算量大,而且有些区域样品的光谱信息很弱,与样品的组成或性质间缺乏相关关系。所以通过主成分分析,选取对于机油品种敏感的新变量来建立品种鉴别模型。可大大简化模型,提高计算速度和分类准确度。表 1 表示前 8 个主成分对原始变量的解释程度。

表 1 前 8 个主成分及其累积贡献率

Table 1 First eight principal components and their total explained variae rates

Principal component	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Accumulative reliabilities /%	52.38	75.63	86.85	89.98	91.87	93.17	94.40	95.38

累积可信度 (Accumulative reliabilities),即主成分能够解释原始波长变量的程度。如表 1 可知前 8 个主成分能够解释原始波长变量的 95.38%,说明前 8 个主成分可代表原可见-近红外光谱的主要信息,说明主成分分析是一种非常有效的数据挖掘方法,它把原来的 600 个波长变量压缩成了彼此正交的新变量,这 8 个新变量彼此间是互不影响的,而且

能代表绝大部分原变量包含的信息。

从三种机油的 180 个样品中各随机抽取 50 个共 150 个样本作为建模集,其余各 10 个样本作为预测集,进行判别分析。对于建模集和预测集中的光谱数据,先利用 Unscrambler 软件获取了光谱数据的前 8 个主成分,选择经典的多类别判别分析法,借助 DPS 软件进行判别分析,给出判别函数,建立多类

判别分析品种鉴别模型。以下是三种机油的判别方程:

壳牌机油:

$$Y_1 = -30.734 + 3851.213x_1 + 1700.889x_2 - 898.221x_3 + 3767.616x_4 + 7080.834x_5 + 6138.231x_6 + 3100.931x_7 + 1291.52x_8; (1)$$

长城牌机油:

$$Y_2 = -7.053 - 1437.6x_1 - 38.8619x_2 - 210.5343x_3 - 1919.43x_4 - 3655.57x_5 - 1766.59x_6 - 2079.52x_7 - 1773.83x_8; (2)$$

玉柴机油:

$$Y_3 = -9.8321 - 2041.95x_1 - 1329.99x_2 + 541.1353x_3 - 1286.32x_4 - 2829.92x_5 - 3084x_6 - 707.631x_7 + 382.1324x_8; (3)$$

表 2 多类判别分析的回代验证结果

Table 2 Resubstitution rate using multiple discriminant analysis

Actual varieties	Varieties discriminated by model			Total	Correct rate /%
	Shell oil	Great Wall oil	Yuchai oil		
Shell oil	50	0	0	50	100
Great Wall oil	0	50	0	50	100
Yuchai oil	0	6	44	50	88

表 3 多类判别分析模型的验证结果

Table 3 Predicting rate of unknown test samples by using multiple discriminant analysis

No.	Discriminated by model	Posterior probability	No.	Discriminated by model	Posterior probability
1	1	100%	16	2	100%
2	1	100%	17	2	100%
3	1	100%	18	2	100%
4	1	100%	19	2	100%
5	1	100%	20	2	100%
6	1	100%	21	3	99.9%
7	1	100%	22	3	100%
8	1	100%	23	3	100%
9	1	100%	24	3	100%
10	1	100%	25	3	100%
11	2	100%	26	3	100%
12	2	100%	27	3	100%
13	2	100%	28	3	100%
14	2	100%	29	3	100%
15	2	100%	30	3	100%

Note: 1—10, Shell oil; 11—20, Great Wall oil; 21—30, Yuchai oil; 1— Shell oil ; 2— Great Wall oil ; 3— Yuchai oil

为了检验所建立的判别分析模型的可靠程度,对未参与建立模型的 30 个样本进行预测验证。30 个样本依次是壳牌机油 10 个,长城牌机油 10 个,玉

柴牌机油 10 个。将 30 个样本的相关数据分别代入三个判别方程中,得到表 3。从表 3 中我们可以看到 30 个未参加建模的样本,前 10 个被判别值为 1,即是壳牌机油;第 11 个到第 20 被判别值为 2,即是长城牌机油;最后 10 个被判别值为 3,即是玉柴牌机油。模型的判别品种与实际品种完全一致,除了第 21 号样本的后验概率是 99.9%,其它样本的后验概率都是 100%,该模型对于未知样本预测正确率达到 100%。

将已知 150 个样本代入已建立的判别方程,按各母体的后验概率重新归类。如果重新归类结果与已知类别的符合率很高,则判别方程的判别效果就好。其判别结果见表 2 所示。从表中可以看出,所建立的判别模型对壳牌机油、长城牌机油和玉柴牌机油的判识效果均相当显著,正确率分别为 100%、100% 和 88%。从总体上看,在将已知样本代入判别函数进行回判时,其总的回判率达 96%,说明所建立的判别模型是可靠的。

柴牌机油 10 个。将 30 个样本的相关数据分别代入三个判别方程中,得到表 3。从表 3 中我们可以看到 30 个未参加建模的样本,前 10 个被判别值为 1,即是壳牌机油;第 11 个到第 20 被判别值为 2,即是长城牌机油;最后 10 个被判别值为 3,即是玉柴牌机油。模型的判别品种与实际品种完全一致,除了第 21 号样本的后验概率是 99.9%,其它样本的后验概率都是 100%,该模型对于未知样本预测正确率达到 100%。

## 4 结 论

本文提出了将主成分分析与多类判别分析相结合的方法进行光谱数据的分析和建模,建立的机油品种判别模型能有效地从大量光谱信息中提取有用信息,降低数据维数,简化运算,较好地快速判别机油的品种。模型对三种机油品种的回判率达 96%,预测未知样本的机油品种识别率达 100%。说明运用可见-近红外光谱技术可以快速、准确的对机油品种进行鉴别,具有很强的实用性,为其他产品的品种鉴别分析提供了一种新的方法。

## 参 考 文 献

1 Zhang Lingfan, Guo Jianwei. Fuzzy comprehensive evaluation of

- wear status on a type of engine group [J]. *Lubrication Engineering*, 2007, **32**(10): 120~122
- 张凌芳, 郭建维. 某型发动机组的磨损状况的模糊综合评判[J]. 润滑与密封, 2007, **32**(10): 120~122
- 2 Li Jianjun. Effect of engine oil property on environmental protection, energy saving and wear resistance [J]. *Diesel Engine*, 2003, **2**: 36~39
- 李建军. 机油品质对环保、节能和机械磨损的影响[J]. 柴油机, 2003, **2**: 36~39
- 3 Wei Shaoyuan. Analysis of the effects of the engine lubricant quality on the wear to the engines [J]. *Lubrication Engineering*, 2002, **2**: 88~90
- 卫绍元. 发动机润滑油品质对发动机磨损影响的分析[J]. 润滑与密封, 2002, **2**: 88~90
- 4 Cai Ruibin, Wang Bin, Zhu Yuqiang. Experimental research and technological economic analysis of a new oil lubricant nursing element [J]. *Machine Tools & Hydraulics*, 2008, **36**(4): 77~80
- 蔡锐彬, 王斌, 朱玉强. 新型机油添加剂的实验研究与技术经济分析[J]. 机床与液压, 2008, **36**(4): 77~80
- 5 Guan Liang, Feng Xinlu, Xiong Gang *et al.*. Rapid determination of the viscosity of automobile lube oil by dielectric spectroscopy [J]. *Petroleum Processing and Petrochemicals*, 2008, **39**(2): 64~68
- 管亮, 冯新泸, 熊刚等. 介电谱技术快速测定车用润滑油粘度[J]. 石油炼制与化工, 2008, **39**(2): 64~68
- 6 Han Donghai, Wang Jiahua. Review of nondestructive measurement of fruit quality by means of near infrared spectroscopy[J]. *Chinese J. Lasers*, 2008, **35**(8): 1123~1131
- 韩东海, 王加华. 水果内部品质近红外光谱无损检测研究进展[J]. 中国激光, 2008, **35**(8): 1123~1131
- 7 Zhao Jiewen, Guo Zhiming, Chen Quansheng *et al.*. Feasibility study on use of near-infrared spectroscopy in quantitative analysis of catechins in green tea[J]. *Acta Optica Sinica*, 2008, **28**(12): 2302~2306
- 赵杰文, 郭志明, 陈全胜等. 近红外光谱法快速检测绿茶中儿茶素的含量[J]. 光学学报, 2008, **28**(12): 2302~2306
- 8 Wang Li, He Ying, Wang Yanping. Oil spill identification by near-infrared spectroscopy coupled with clustering analysis based on principle components [J]. *Marine Environmental Science*, 2004, **23**(2): 58~61
- 王丽, 何鹰, 王颜萍等. 近红外光谱技术结合主成分聚类分析判别海面溢油种类[J]. 海洋环境科学, 2004, **23**(2): 58~61
- 9 Jiang Lulu, Tan Lihong, Qiu Zhengjun *et al.*. Discrimination of varieties of brake fluid using visual-near infrared spectra [J]. *Spectroscopy and Spectral Analysis*, 2008, **28**(6): 1296~1299
- 蒋璐璐, 谈黎虹, 裘正军等. 基于可见-近红外光谱的制动液品牌鉴别方法研究[J]. 光谱学与光谱分析, 2008, **28**(6): 1296~1299
- 10 Tian Gaoyou, Xiong Chunhua, Liu Huiying. Model recognition of gasoline by NIR technology [J]. *Modern Scientific Instruments*, 2006, (5): 84~87
- 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, (5): 84~87
- 11 Liu Fei, He Yong, Wang Li. Methods for the prediction of sugar content of rice wine using visible-near infrared spectroscopy[J]. *Acta Optica Sinica*, 2007, **27**(11): 2054~2058
- 刘飞, 何勇, 王莉. 黄酒糖度预测的可见-近红外光谱方法研究[J]. 光学学报, 2007, **27**(11): 2054~2058
- 12 Chen Xiaojing, Wu Di, Yu Jijia *et al.*. A new choice method of characteristic wavelength of visible/near infrared spectroscopy [J]. *Acta Optica Sinica*, 2008, **28**(11): 2153~2158
- 陈孝敬, 吴迪, 虞佳佳等. 一种用于可见-近红外光谱特征波长选择的新方法[J]. 光学学报, 2008, **28**(11): 2153~2158
- 13 Zhang Qianqian, Lei Shuhe, Wang Xiulin *et al.*. Research on discrimination of 3D fluorescence spectra of phytoplanktons [J]. *Spectroscopy and Spectral Analysis*, 2004, **24**(10): 1227~1229
- 张前前, 类淑河, 王修林等. 浮游植物活体三维荧光光谱分类判别方法研究[J]. 光谱学与光谱分析, 2004, **24**(10): 1227~1229
- 14 Wang Haishui, Wang Dongmei, Xi Shixuan. The application of near infrared spectroscopy for qualitative and quantitative analysis [J]. *Analysis and Testing Technology Instruments*, 2002, **8**(3): 136~138
- 王海海水, 汪冬梅, 席时权. 近红外光谱在品质分析和定量分析中的应用[J]. 分析测试技术与仪器, 2002, **8**(3): 136~138
- 15 Xun Lina, Fang Yonghua, Li Xin. A small target detection approach based on endmember extraction in hyperspectral image [J]. *Acta Optica Sinica*, 2007, **27**(7): 1178~1182
- 寻丽娜, 方勇华, 李新. 高光谱图像中基于端元提取的小目标检测算法[J]. 光学学报, 2007, **27**(7): 1178~1182