

文章编号: 1672-8785(2011)08-0035-04

基于可见-近红外光谱技术的制动液品牌混掺比例快速无损检测方法

蒋璐璐¹ 石 慧² 吴 迪² 魏 萱² 谈黎虹¹ 何 勇² 朱 枫¹

(1. 浙江经济职业技术学院, 浙江 杭州 310018;

2. 浙江大学生物系统工程与食品科学学院, 浙江 杭州 310058)

摘 要: 研究了基于可见-近红外光谱技术的制动液品牌混掺比例快速无损检测方法。全波段建立的偏最小二乘回归 (PLSR) 和最小二乘支持向量机 (LS-SVM) 模型均得到了较好的预测结果。这两个模型的建模集和预测集的确定系数 (r_c^2 和 r_p^2) 均在 0.98 以上。采用连续投影算法 (SPA) 挖掘特征波长, 最终选择了 439 nm、443 nm、459 nm、519 nm、570 nm、717 nm、896 nm 和 902 nm 共 8 个波长作为最优变量。基于 SPA 选择的变量建立的 PLSR 和 LS-SVM 模型的 r_c^2 和 r_p^2 均在 0.97 以上, 能够满足实际应用的需要。研究结果表明, 可见-近红外光谱可以用于制动液品牌混掺比例快速无损检测。

关键词: 可见-近红外光谱; 制动液; 品牌混掺; 偏最小二乘回归; 最小二乘支持向量机; 连续投影算法

中图分类号: TK418; O657.32 **文献标识码:** A DOI: 10.3969/j.issn.1672-8785.2011.08.007

Fast and Non-invasive Measurement of Different Mixture Percentages of Brake Fluid Based on Visible and Near Infrared Spectroscopy

JIANG Lu-lu¹, SHI Hui², WU Di², WEI Xuan², TAN Li-hong¹, HE Yong², ZHU Feng¹

(1. Zhejiang Technology Institute of Economy, Hangzhou 310018, China;

2. College of Biosystems Engineering & Food Science, Zhejiang University, Hangzhou 310058, China)

Abstract: A fast and non-invasive method for determining the mixture percentages of brake fluid based on visible and near infrared spectroscopy (Vis-NIRS) was proposed. Both a partial least square regression (PLSR) model and a least-square support vector machine (LS-SVM) model were established according to the spectra obtained in the whole wavelength range. With those two models, good prediction results were obtained. The determination coefficients of their calibration and prediction sets (r_c^2 and r_p^2) were greater than 0.98. The successive projection algorithm (SPA) was used to select the effective variables. Finally, eight variables of 439 nm, 443 nm, 459 nm, 519 nm, 570 nm, 717 nm, 896 nm and 902 nm were selected as the optimal variables to be input into the PLSR and LS-SVM models. The r_c^2 and r_p^2 of both two models were greater than 0.97 which was adequate for practical application. It was concluded that Vis-NIRS could be used to fastly and non-invasively determine the mixture percentages of brake fluid.

Key words: visible and near infrared spectroscopy; brake fluid; mixture; partial least square regression; least-square support vector machine; successive projection algorithm

收稿日期: 2011-07-10

基金项目: 国家“十一五”科技支撑计划项目 (2006BAD10A0403); 浙江省教育厅科技项目 (20071275)

作者简介: 蒋璐璐 (1972-), 女, 浙江金华人, 副教授, 主要从事汽车检测方面的教学与科研工作。

E-mail: abbyu111@sina.com

0 引言

制动液的质量优劣、品牌选择以及使用正误等因素将会直接影响机械对象的运行安全。不同品牌和类型制动液的混掺,由于各自配方的不同,将会导致混合制动液结构分层和内部成分发生化学变化,从而影响制动液的性能,降低汽车等的制动效果,甚至会产生制动失灵等严重后果^[1]。对于不同品牌的同类制动液,即使相溶性较好也不能混用^[2]。由于不同品牌和品质的制动液之间存在较大的价格和性能差异,一些不法商家会将不同品质的制动液混掺后进行销售以牟取暴利。常用的制动液混掺比例检测不仅耗时费力、成本高、程序繁琐,而且还需要专业的技术人员来执行。因此,人们需要一种可以快速、无损检测制动液混掺比例的方法。

可见-近红外光谱法是一种快速、简便、非破坏性分析方法。它已被广泛应用于食品、石油化工和医药等行业^[3]。在机油检测领域,光谱技术已应用于润滑油测试及其成分的定量分析与研究^[4-8]。但目前对制动液品牌混掺比例进行检测的研究还没有相关报道。本文采用可见-近红外光谱技术对制动液混掺比例进行检测。在建立全波段光谱数据的基础上,分别采用主成分分析法(PCA)和连续投影算法(SPA)两种方法来减少模型的输入变量个数。同时分别采用偏最小二乘回归(PLSR)和最小二乘支持向量机(LS-SVM)建立光谱检测模型,最终实现对制动液混掺比例的快速、无损检测。

1 材料与方 法

1.1 仪器设备

实验使用美国 ASD 公司生产的 FieldSpec Handheld 光谱仪采集不同混掺比例制动液的可见-近红外光谱。光谱仪的检测波长范围为 325 ~ 1075 nm。检测光源采用日本 Ushio Lighting 公司生产的 14.5 V 卤素灯。此外,采用 Unscramble V9.7 和 MATLAB 7.6 软件进行化学计量学分析。

1.2 样品准备和光谱采集

试验选用灵峰和 TRWDOT4 牌制动液进行互掺,分别在 TRWDOT4 牌制动液中添加 10 %、20 %、40 % 和 60 % 4 种不同比例的灵峰牌制动液。每种类型获取 40 个样本,共计 160 个样本。采用透射法进行光谱扫描,制动液的扫描光程为 5 mm。每个样本扫描 30 次,保存 3 条光谱,计算平均光谱后将其作为该样本的最终光谱。图 1 所示为不同混掺比例制动液的可见-近红外光谱的吸光度值。从 4 种不同混掺量的制动液中各自随机选择 20 个样本,并将这 80 个样本作为建模集,同时将每种类型剩余的 20 个样本(共 80 个样本)作为预测集。

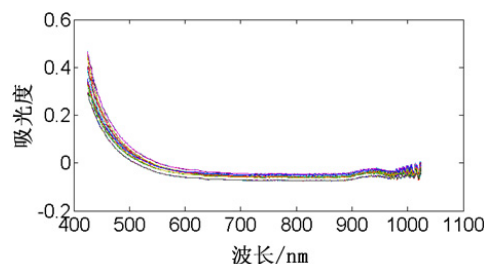


图 1 不同混掺比例制动液的可见-近红外光谱的吸光度值

1.3 光谱化学计量学分析

由于实验获取的光谱数据在波长范围两端存在一定噪声,选取 425 ~ 1024 nm 波段(共 600 个变量)的光谱数据进行分析。将测得的透射光谱数据转换成吸光度值后,选用 Savitzky-Golay 平滑法结合变量标准化(SNV)进行光谱预处理。

主成分分析法(PCA)是化学计量学中一种常用的线性映射方法,它广泛运用于光谱数据降维、聚类分析、变量提取以及与其他计量学算法联用等方面^[9]。连续投影算法(SPA)是一种新提出的变量选择算法。它通过对向量进行投影,挖掘相互间冗余度较小的变量组合,最终获得共线性较小的变量^[10]。

偏最小二乘法回归(PLSR)是一种经典的光谱建模算法。它通过对变量 X 和 Y 进行分解,同时提取成分(通常称为因子),从而得到一组最小误差平方的最佳函数匹配^[6]。最小二乘支持向量机(LS-SVM)则是一种常用的光谱非线性建模

算法。该算法采用非线性映射函数将输入变量映射到高维特征空间,从而将优化问题改成等式约束条件。同时它基于拉格朗日乘子对每个变量进行偏微分求解,并以求解一组线性方程代替经典 SVM 中复杂的二次优化问题来获取支持向量,从而降低计算复杂性,并加快计算速度^[11]。

本文建立的光谱模型采用以下参数进行评估:建模集和预测集的确定系数(r_c^2 和 r_p^2)、校正建模均方根误差(RMSEC)和预测均方根误差(RMSEP)。好的模型需要有较高的 r^2 和较低的均方根误差。

2 试验结果与分析

2.1 全波段光谱模型分析

基于预处理后的全波段光谱数据,分别采用 PLSR 和 LS-SVM 建立制动液品牌混掺比例快速检测模型。表 1 列出了该模型的建模和预测结果(见 1~2 行)。从表 1 中可以看出,LS-SVM 模型和 PLSR 模型的 r_c^2 和 r_p^2 均在 0.98 以上。结果表明,利用可见-近红外光谱技术能够实现制动液品牌混掺比例的快速无损检测。同时通过对比可知,LS-SVM 模型的结果比 PLSR 模型的好。其中,LS-SVM 模型的 RMSEC 值仅为 0.024,比 PLSR 模型的减少了 98.80%。而 LS-SVM 模型的 RMSEP 值仅为 0.517,比 PLSR 模型的减少

了 76.51%。由于 LS-SVM 为非线性建模算法,而 PLSR 为线性建模算法,这就说明可见-近红外光谱和制动液品牌混掺比例之间的定量关系能够通过非线性算法进行更好的表达。

2.2 主成分分析

当采用全波段光谱建立光谱模型时,由于共有 600 个输入变量,光谱模型将会过于冗余,从而影响建模速率和预测精度。同时光谱信息中还夹杂着噪声等干扰因素,这也会对光谱建模造成影响。因此,本文在光谱数据建模前先对 600 个原始光谱变量进行 PCA 分析。表 2 列出了主成分的累计贡献率。其中,第一个主成分的贡献率已经达到了 68.74%,说明第一个主成分所包含的信息能够解释六成以上的全光谱信息。当主成分的个数达到 8 个时,主成分的累计贡献率已经达到了 99% 以上。因此,本文采用前 8 个主成分代替全光谱数据作为输入变量,分别建立 PLSR 和 LS-SVM 模型。表 1 列出了该模型的建模和预测结果(见 3~4 行)。从表 1 中可以看出,虽然输入变量从 600 个减少到了 8 个,但是模型的整体预测结果却没有下降。PCA-LS-SVM 模型和 PCA-PLSR 模型的 r_c^2 和 r_p^2 依然大于 0.98,说明前 8 个主成分能够代替全光谱数据进行制动液品牌混掺比例的快速检测。通过对比可知,LS-SVM 模型的结果依然比 PLSR 模型的好。其

表 1 制动液品牌混掺比例快速检测模型的各项结果

变量选择	变量个数	建模	LVs	R_c^2	RMSEC	R_p^2	RMSEP
全波段	600	PLSR	3	0.989	1.973	0.987	2.199
全波段	600	LS-SVM	/	1.000	0.024	0.999	0.517
PCA	8	PLSR	3	0.989	2.218	0.987	2.204
PCA	8	LS-SVM	/	0.999	0.410	0.998	0.726
SPA	8	PLSR	4	0.986	2.217	0.975	3.031
SPA	8	LS-SVM	/	1.000	0.119	0.997	0.953

LVs: Latent variables, 潜在变量

表 2 进行主成分分析后,前 10 个主成分的累计贡献率

主成分	贡献率 / %	累计贡献率 / %	主成分	贡献率 / %	累计贡献率 / %
PC1	68.74	68.74	PC6	2.23	97.47
PC2	9.80	78.54	PC7	1.26	98.73
PC3	9.40	87.95	PC8	0.59	99.32
PC4	4.39	92.33	PC9	0.19	99.51
PC5	2.91	95.24	PC10	0.08	99.59

中, LS-SVM 模型的 RMSEC 值仅为 0.410, 比 PLSR 模型的下降了 81.53%。LS-SVM 模型的 RMSEP 值仅为 0.726, 比 PLSR 模型的下降了 67.04%。这说明非线性算法能够更好地描述主成分和制动液品牌混掺比例之间的定量关系。

2.3 连续投影算法

在主成分分析中, 为了得到主成分, 必须获得全波段的光谱数据, 然后才能进行主成分分析。由于全波段光谱的测量需要价格昂贵的测试仪器, 在制动液品牌混掺比例快速检测中无法实现光谱技术快速、低成本的特点。因此, 在光谱分析中最好能够获得若干个对于模型建立比较重要的光谱特征变量, 从而为开发基于特征波长的简易、低成本的光谱仪器提供理论依据。

本文采用 SPA 对全波段光谱共 600 个变量进行有效变量选择。当通过交互验证得到有效变量个数为 8 个时, 模型的验证标准偏差达到最优。图 2 为采用 SPA 选择的 8 个波长在全波段范围内的分布图。其中, 黑色竖线对应选中的波长 (分别为 439 nm、443 nm、459 nm、519 nm、570 nm、717 nm、896 nm 和 902 nm)。将得到的 8 个变量作为输入, 分别建立 PLSR 和 LS-SVM 模型 (结果见表 1 的 5~6 行)。可以看出, SPA 选择后的变量个数和全波段光谱相比减少了 98.67%, 而模型精度并没有明显下降。这说明选取得到的 8 个特征波长能够代替全光谱波长进行制动液品牌混掺比例的快速检测。

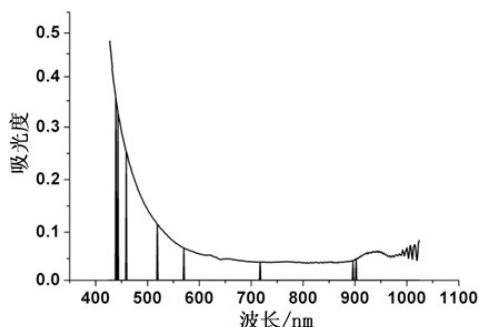


图 2 采用 SPA 选择的 8 个波长在全波段范围内的分布图

通过对比可知, LS-SVM 模型的结果依然比 PLSR 模型的好。其中, LS-SVM 模型的 RMSEC 值仅为 0.119, 比 PLSR 模型的下降了 94.63%。LS-SVM 模型的 RMSEP 值仅为 0.953, 比 PLSR

模型的下降了 68.57%。这说明非线性算法能够更好地描述 SPA 选择的特征波长和制动液品牌混掺比例之间的定量关系。

3 结束语

本文利用可见-近红外光谱技术对制动液品牌混掺比例快速无损检测方法进行了研究。全波段建立的 PLSR 和 LS-SVM 模型均得到了较好的预测结果。两个模型的 r_c^2 和 r_p^2 均在 0.98 以上。通过 SPA 计算最终选择 8 个波长作为最优变量, 变量个数和全波段光谱相比减少了 98.67%, 而模型精度并没有明显下降。这说明选取得到的 8 个特征波长能够代替全光谱波长进行制动液品牌混掺比例的快速检测。SPA-PLSR 模型和 SPA-LS-SVM 模型的 r_c^2 和 r_p^2 均在 0.97 以上, 能够满足实际应用的需要。

参考文献

- [1] 张传爱, 周朝霞. 注意制动液的选择与使用 [J]. 建筑机械化, 2001, 22(3): 47-49.
- [2] 石俊峰, 褚丽, 刘宏, 等. 合成制动液混用问题研究 [J]. 石油商技, 1999, 17(6): 23-25.
- [3] 严衍禄, 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.
- [4] 张瑜, 吴迪, 蒋璐璐, 等. 发动机润滑油品质的可见-近红外光谱快速检测 [J]. 光谱实验室, 2010, 27(4): 1629-1632.
- [5] 蒋璐璐, 邵咏妮, 张瑜, 等. 基于光谱技术和连续投影算法的润滑油品牌快速鉴别研究 [J]. 光谱实验室, 2010, 27(4): 1507-1509.
- [6] 张瑜, 蒋璐璐, 吴迪, 等. 发动机润滑油品质的可见-近红外光谱快速检测 [J]. 光谱学与光谱分析, 2010, 30(8): 1629-1632.
- [7] 赵芸, 蒋璐璐, 张瑜, 等. 应用光谱技术快速测定发动机润滑油的黏度值 [J]. 光谱学与光谱分析, 2010, 30(9): 2496-2499.
- [8] 马兰芝, 褚小立, 田松柏, 等. 红外光谱法在润滑油分析中的应用与研究进展 [J]. 分析仪器, 2010, 28(2): 1-4.
- [9] 吴迪, 金春华, 何勇, 等. 基于连续投影算法的光谱主成分组合优化方法研究 [J]. 光谱学与光谱分析, 2009, 29(10): 2734-2737.
- [10] 吴迪, 曹芳, 张浩, 等. 基于可见-近红外光谱技术的水稻穗颈瘟染病程度分级方法研究 [J]. 光谱学与光谱分析, 2009, 29(12): 3295-3299.
- [11] Wu D, Yang H Q, Chen X J, et al. Application of Image Texture for The Sorting of Tea Categories Using Multi-spectral Imaging Technique and Support Vector Machine [J]. Journal of Food Engineering, 2008, 88(4): 474-483.