

文章编号:1672-058X(2012)03-0094-04

# 基于近红外光谱的透平油微量水分检测\*

侯明明,喻其炳,焦昭杰,李川\*\*

(重庆工商大学 废油资源化技术与装备教育部工程研究中心,重庆 400067)

**摘要:**传统的石油产品水分检测存在试剂有毒、操作复杂、测量时间长、受环境影响大、满足不了在线快速检测等问题;在此利用近红外光谱快速测定透平油中微量水分,利用关联向量机对测定的光谱数据进行建模,实验结果表明:最大绝对误差和均方根误差分别为  $20.99 \times 10^{-6} (V/V)$ 、 $11.11 \times 10^{-6} (V/V)$ ,说明该模型具有较高的预测能力,对实现透平油微量水分的快速检测具有重要意义。

**关键词:**透平油;近红外光谱;微量水分;关联向量机

**中图分类号:**X705

**文献标志码:**A

近红外区域是指波长在 780 ~ 2526 nm 的电磁波,近红外光谱主要是由于分子振动的非谐性使分子振动从基态向高能级跃迁时产生的,其光谱信息来源于分子的 C-H、N-H、O-H 等含氢基团发生的倍频及合频对光能量的吸收,不同基团产生的光谱在吸收峰位和强度上有所不同。由于物质在该谱区的倍频和合频吸收信号弱,谱带重叠,解析复杂,给测量带来了一定的难度。随着化学计量学以及计算机技术的发展,这个问题已经得到了很好地解决,近红外光谱技术成为了 20 世纪 90 年代以来发展最快、最引人注目的光谱分析技术之一,它具有分析速度快、操作简单、所需样品少、无需预处理、无污染无损检测等特点,被誉为分析的巨人。已被广泛应用于石油、化工、农产品、制药、和食品检测等行业<sup>[1-2]</sup>。

水分含量是透平油的关键性指标之一。油中水分的存在会破坏透平油形成的油膜,使油品的润滑效果变差,加速透平油各项性能的劣化,降低设备的使用寿命,对设备的正常运行构成严重威胁<sup>[3]</sup>。常规的油品水分检测方法有蒸馏法、色谱法、卡尔-费休法、重量法等<sup>[4]</sup>。由于操作复杂,用时长,受环境影响大,副反应多,试剂有毒,精度不高等原因,这些传统方法不能满足大多数用户要求在线快速检测油品水分的需求。本文正是基于这一现状,采用现代近红外光谱技术,结合关联向量机建模的方法对透平油中微量水分进行了检测,旨在提出一种快速无损的,适合快速检测的透平油中微量水分的定量分析方法。

收稿日期:2010-06-15;修回日期:2011-09-20.

\* 基金项目:国家自然科学基金(50905193);重庆市科技计划项目(2009DA0001-A13,2010BB4249,2010BB4261).

作者简介:侯明明(1983-),男,陕西延安人,硕士研究生,从事利用近红外光谱分析检测油品中微水含量方面的研究.

\*\* 通讯作者:李川(1975-),男,博士,重庆市人,副研究员,从事先进检测技术研究,E-mail:chuanli@21cn.com.

## 1 实验内容

### 1.1 实验仪器

NIRQuest-512 近红外光谱仪(美国 Ocean Optics 公司生产),InGaAs 检测器,波长重复性  $\pm 0.05$  nm,光学分辨率 4 nm,波长范围 900 ~ 1 700 nm。

### 1.2 实验方法

(1) 样品的配置。取未开封的 46# 透平油,用移液管移取 10 mL 油品于 25 mL 的试管中,用 0.1  $\mu\text{L}$  的微量注射器向其中注射蒸馏水配置成 40 个不同浓度梯度的含油水样,再经过超声波处理 40 min 使油水混合均匀,作为实验样品。实验样品的含水量梯度分布较均匀,总体含水量分布范围在  $0 \sim 400 \times 10^{-6}$  (V/V) 之间,分布范围较广,具有一定的代表性。

(2) 光谱数据的采集。光谱采样积分时间选择 720 ms,对光谱进行 2 度平滑,采用 InGaAs 检测器。测定样品的吸光度,40 个不同样品的近红外光谱如图 1。

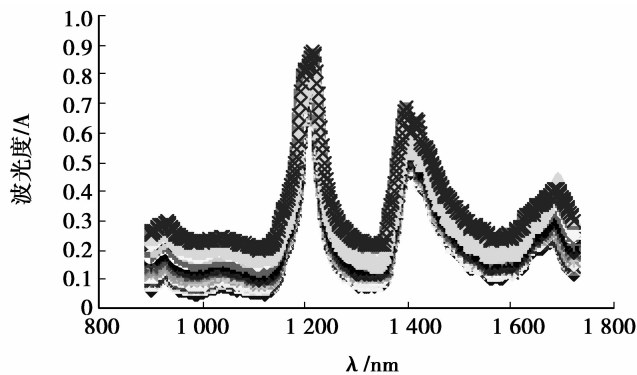


图 1 样品的近红外光谱

## 2 关联向量机建模

### 2.1 关联向量机的基本理论

关联向量机(Relevance Vector Machines, RVM)是 M. T. Tipping 提出的一种与支持向量机相似的稀疏概率模型<sup>[5]</sup>,它的训练是在贝叶斯框架下进行的,可以用它进行回归估计预测,从而获得预测值的分布。RVM 将拉格朗日系数向量  $w$  作为权参数,建立了相应的显著度框架,由此求得的非零  $w_i$  所对应得学习样本被称为相关向量(Relevance Vector, RV)<sup>[6]</sup>。

给定输入向量  $x = \{x_1, x_2, \dots, x_n, \dots, x_N\}^T$  和相应的输出  $t = \{t_1, t_2, \dots, t_n, \dots, t_N\}^T$ ,学习的目的就是应用这些训练数据和先验知识来设计一个系统,使系统对新的输入  $x$  预测输出  $t$ 。假设目标值是一个未知函数和一些噪声的组合:

$$t = y(x;w) + \varepsilon \quad (1)$$

式(1)中,  $w$ -模型权值,  $w = \{w_1, w_2, \dots, w_i, \dots, w_M\}^T$ ;  $\varepsilon$ -均值为0, 方差为  $\sigma^2$  的噪声,  $\varepsilon = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n, \dots, \varepsilon_N\}^T$ 。  $y(x; w)$  为函数族, 由下式给出:

$$y(x; w) = \sum_{i=1}^M w_i \Phi(x) + w_0 \quad (2)$$

式(2)中,  $\Phi(x)$  为一组非线性基函数, 有:

$$\Phi(x) = \{\phi_1(x), \phi_2(x), \dots, \phi_i(x), \dots, \phi_M(x)\}^T \quad (3)$$

支持向量机要求使用的基函数必须满足 Mercer 核定理, 而 RVM 无此限制, 研究选用每个训练样本为中心的 Gaussian 函数作为基函数。在定义了基函数后, 可以在贝叶斯框架下用极大似然方法来训练权值  $\{w_i\}$ 。式(3)相应的训练样本集的似然函数:

$$p(t | w, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{\|t - w\Phi\|^2}{2\sigma^2}\right\} \quad (4)$$

式(4)中的  $w, \sigma^2$  的极大似然估计一般会导致严重的过学习问题, 因此稀疏贝叶斯学习方法对权值  $w$  赋予先验的条件概率分布:

$$p(w | \alpha) = \prod_{i=0}^N N(w_i | 0, \alpha_i^{-1}) \quad (5)$$

式(5)中,  $\alpha$  为超参数矢量。每个权值对应一个超参数, 最终导致解的稀疏性更好。根据贝叶斯规则, 可以获得在权值上的后验概率:

$$p(w | t, \alpha, \sigma^2) = (2\pi)^{\frac{-(N+1)}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(w - \mu)^T \Sigma^{-1}(w - \mu)\right\} \quad (6)$$

$$\Sigma = (\Phi^T B \Phi + A)^{-1} \quad (7)$$

$$\mu = \Sigma \Phi^T B t \quad (8)$$

式(6)中,  $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$ ;  $B = \sigma^{-2} I_N$ 。  $\sigma^2$  也被当作超参数来处理, 可以从数据中得到估计。

通过对权值积分, 可以得到超参数的边缘似然或称证据函数:

$$p(t | \alpha, \sigma^2) = (2\pi)^{\frac{-N}{2}} |B^{-1} + \Phi A^{-1} \Phi^T|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} t^T (B^{-1} + \Phi A^{-1} \Phi^T)^{-1} t\right\} \quad (9)$$

通过最大化上式的超参数似然分布求  $\alpha, \sigma^2$  的可能值, 由于不能得到  $\alpha_{MP}, \sigma_{MP}^2$  的解析形式, 故采用反复迭代估计法。对上式求导并重排得:

$$\alpha_i^{\text{new}} = \gamma_i / \mu_i^2 \quad (10)$$

式(10)中,  $\mu_i$  根据式(8)计算得到的第  $i$  个后验权值平均值;  $\gamma_i = 1 - \alpha_i \Sigma_{ii}$ ,  $\Sigma_{ii}$  是用当前的  $\alpha, \sigma^2$  根据式(7)计算得后验权值协方差矩阵的第  $i$  个对角线元素。同样对噪声方差  $\sigma^2$  求导, 得到相同的估计:

$$(\sigma^2)^{\text{new}} = \|t - \Phi \mu\|^2 / (N - \sum_i \gamma_i) \quad (11)$$

学习算法反复计算式(10)和式(11), 同时更新式(7)和式(8), 直到满足合适的收敛条件。在迭代过程中, 有许多  $\alpha_i \rightarrow \infty$ , 可以删除这一部分权值和基函数, 从而保证了解的稀疏性。当上述超参数估计过程收敛, 则根据权值的后验及最优超参数  $\alpha_{MP}, \sigma_{MP}^2$ , 可以对新的观测数据  $x^*$  进行回归估计, 计算其预测分布:

$$p(t^* | t, \alpha_{MP}, \sigma_{MP}^2) = N(\mu^*, \sigma_*^2) \quad (12)$$

式(12)中,  $\sigma_*^2 = \sigma_{MP}^2 + \varphi(x^*)^T \Sigma \varphi(x^*)$ 。因此, 可以直观的得到预测的均值, 也就是非线性回归估计:

$$y(x^*; w) = \mu^* = \mu^T \varphi(x^*) \quad (13)$$

RVM 使用贝叶斯框架下的显著度解决了模型的参数选取问题,具有较好的适用性。基于 RVM 的非线性回归能力,已有研究将其应用到软测量建模过程的非线性函数映射<sup>[7]</sup>。

## 2.2 模型建立

在总共的 40 个实验样品中,用 25 个样品作为训练集,15 个样品作为校验集。采用 RVM 法对全波范围内建立油样的含水量定量分析模型,定义均方根误差  $RMSE$  和最大绝对误差  $MAXAE$  来衡量模型的精度<sup>[7]</sup>。

$$\begin{cases} RMSE = \sqrt{\frac{\sum_{i=1}^l (y_i - \hat{y}_i)^2}{l}} \\ MAXAE = \max_{i=1}^l |y_i - \hat{y}_i| \end{cases} \quad (14)$$

式(14)中, $l$ 为测试样本的长度; $\hat{y}_i$ 为模型输出预测值; $y_i$ 为实际值。测试结果如图2所示。测得的  $MAXAE$  为  $20.99 \times 10^{-6} (V/V)$ ,  $RMSE$  为  $11.11 \times 10^{-6} (V/V)$ 。

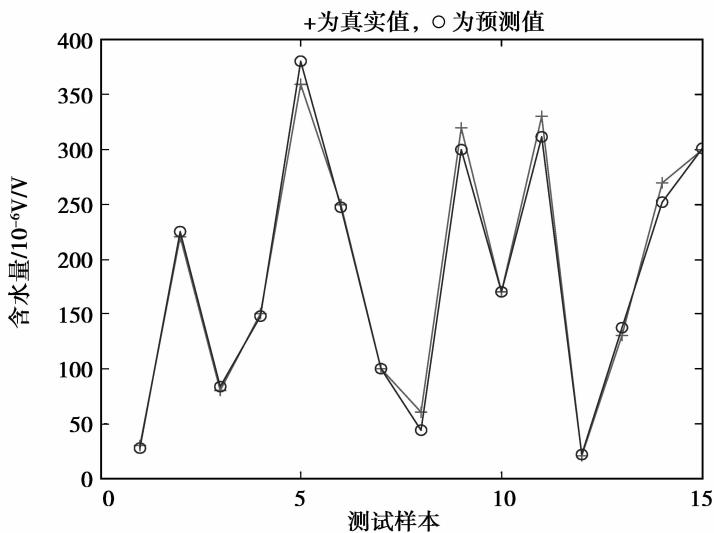


图2 RVM模型的预测值与真实值比较

## 3 结论

(1) 由于水分在近红外光谱区的特征吸收特性,可以利用近红外方法快速检测透平油中微量水分含量,这是对透平油中水分检测技术的一个补充,为油品的在线快速无损检测提供了有力的技术支持。

(2) 计算所得的  $MAXAE$  和  $RMSE$  分别为  $20.99 \times 10^{-6} (V/V)$ 、 $11.11 \times 10^{-6} (V/V)$ ,相较于总体含水量分布范围( $0 \sim 400 \times 10^{-6} V/V$ )表明提出的 RVM 建模方法具有简单的非线性建模结构和明确的物理意义,有较高准确的预测能力。

**参考文献:**

- [1] 高荣强,范世福. 现代近红外光谱分析技术的原理及应用[J]. 分析仪器,2002,10(3):9-12
- [2] 徐广通,袁洪福,陆婉珍. 现代近红外光谱技术及应用进展[J]. 光谱学与光谱分析,2000,20(2):134-142
- [3] 方利民,林敏. 柴油近红外光谱的独立分量分析方法[J]. 石油学报(石油加工),2008,24(6):726-732
- [4] 洪镇南,春日正男,李祖林,等. 基于自适应模型神经网络的变压器油中水分含量检测的研究[J]. 变压器,2008,45(5):34-37
- [5] TIPPING M T. Sparse Bayesian learning and the relevance vector machine[J]. Journal of Machine Learning Research (S1532-4435),2001,1(3):211-244
- [6] 刘遵雄,张德运,孙钦东,等. 基于相关向量机的电力负荷中期预测[J]. 西安交通大学学报,38(10):1006-1008
- [7] 李川,王时龙,张贤明. 基于多关联向量机的动态软测量建模[J]. 系统仿真学报,2009,21(12):3513-3517

## Near Infrared Spectroscopy Measurement of Moisture Content in Turbine Oil Using Relevance Vector Machines

**HOU Ming-ming, YU Qi-bing, JIAO Zhao-jie, LI Chuan**

(Engineering Research Center for Waste Oil Recovery Technology and Equipment of Ministry of Education,  
Chongqing Technology and Business University, Chongqing 400067, China)

**Abstract:** There are evident limitations in traditional measurement of moisture content in oils, such as poisonous reagents, complicated operations, time-consuming, susceptible to environmental interferences and unable on-line quick measurement. A fast determination of the moisture content in turbine oil by near infrared spectroscopy has been introduced in this paper. The relevance vector machine was employed to model the spectral data. The Maximum Absolute Error and the Root Mean Square Error of the experimental results are respectively  $20.99 \times 10^{-6}$  (V/V) and  $11.11 \times 10^{-6}$  (V/V). The results indicate that the proposed model features good prediction performance and thus plays an important role in fast measurement of moisture content in turbine oil.

**Key words:** turbine oil; near infrared spectroscopy; moisture content; relevance vector machine

责任编辑:代小红