

文章编号: 0253-2239(2008)11-2153-06

一种用于可见-近红外光谱特征波长选择的新方法

陈孝敬¹ 吴迪² 虞佳佳² 何勇² 刘守¹

(¹ 厦门大学物理系, 福建 厦门 361005; ² 浙江大学生物系统工程与食品科学学院, 浙江 杭州 310029)

摘要 提出了一种基于模拟退火(SA)算法和最小二乘法支持向量机(LS-SVM)选择可见-近红外光谱特征波长的新方法(SA-LS-SVM)。该方法用 LS-SVM 作为识别器,用识别率作为 SA 的目标函数,提取合适的特征波长数以及对应的特征波长。3 种不同品牌的润滑油可见-近红外光谱的特征波长分别用 SA-LS-SVM,主成分回归分析(PCA)和偏最小二乘法(PLS)进行处理,提取特征波长或主成分,然后结合反向传播人工神经网络(BP-ANN)对各种处理方法进行识别预测。结果发现,SA-LS-SVM 只需从 751 个数据光谱中提取 4 个特征波长,就可以使三种品牌润滑油的识别率达到了 100%,而其他所有的方法发现预测率都达不到 100%,由此验证了 SA-LS-SVM 的优越性。实验结果表明,SA-LS-SVM 不仅能有效地减少建模的变量数,而且可以提高预测精度。

关键词 可见-近红外光谱分析; 识别模型; 模拟退火算法; 最小二乘法支持向量机

中图分类号 O433 文献标识码 A doi: 10.3788/AOS20082811.2153

A New Choice Method of Characteristic Wavelength of Visible/Near Infrared Spectroscopy

Chen Xiaojing¹ Wu Di² Yu Jiajia² He Yong² Liu Shou¹

¹Department of Physics, Xiamen University, Xiamen, Fujian 361005, China

²College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, Zhejiang 310029, China

Abstract A new method based on simulated annealing algorithm (SA) and least-squares support vector machine (LS-SVM) (SA-LS-SVM) was proposed to select the characteristic wavelength for visible-near infrared (Vis/NIR) spectroscopy discrimination. In order to find suitable numbers of characteristic wavelength and corresponding characteristic wavelength, discriminating rate was used as object function for SA, and LS-SVM was adopted as discrimination model. The Vis/NIR spectroscopy characteristic wavelengths of three categories of lubricant were processed by SA-LS-SVM, principal component analysis (PCA) and partial least squares (PLS) respectively, and then predicted by back-propagation artificial neural network (BP-ANN). The results of experiment showed that discriminating rate by using combination of SA-LS-SVM with BP-ANN reaches 100% only using 4 characteristic wavelengths from total of 751 wavelengths, while discriminating rate did not reach 100% by other methods. The proposed algorithm not only reduced the number of spectral variables, but also improved the discriminating rate.

Key words visible/near infrared spectroscopy; discrimination model; simulated annealing algorithm (SA); least squares-support vector machine (LS-SVM)

1 引 言

可见-近红外光谱技术是近几年迅速发展起来的分析技术,近红外光谱能够反映有机分子中含氢

基团(C-H, N-H)的特征振动信息,其光谱特性与有机物质的类型和含量密切相关,与传统的化学分析方法相比,近红外光谱分析技术具有速度快、效

收稿日期: 2008-02-18; 收到修改稿日期: 2008-05-14

基金项目: 国家自然科学基金项目(30671213)和教育部高等学校优秀青年教师教学科研奖励计划(02411)资助课题。

作者简介: 陈孝敬(1978-),男,博士研究生,主要从事光信息和信号处理等方面的研究。E-mail: chenxj9@163.com

导师简介: 刘守(1948-),男,教授,博士生导师,主要从事光信息与全息处理等方面的研究。

E-mail: liushou@xmu.edu.cn

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、制药工业等领域^[1]。国内外很多学者已经利用近红外光谱技术对物质品种和真伪进行鉴别如茶叶^[2]、香烟^[3]和酸奶^[4]等。虽然光谱技术用于品种识别上有很多的优点,但可见-近红外技术也有其局限性:近红外区的谱带复杂、重叠多、数据量大,这样难以用光谱仪所采集的所有数据来建立预测模型^[5]。当前,对光谱数据进行建模提取光谱主要成分的方法,比较常用的有主成分分析(Principal component analysis, PCA)和偏最小二乘法(Partial least squares, PLS)^[6,7],然而使用 PCA 和 PLS 对光谱数据建立光谱预测模型时,需要确定组成组分的特征波长谱区,减小建模和预测时间,这些都带有主观性^[5]。近几年,由于遗传算法(GA)的普遍使用,也有一些研究者采用这种方法对特征波长进行了选择^[5,8~10],然而谱区的选择依然带有主观性,而且遗传算法往往会早收敛的缺陷,往往寻不到全局最优的点,算法就停止^[11]。

相比传统的建模方法,如 PLS,人工神经网络(ANN)等,LS-SVM 是一种新型的非线性建模方法,它是建立在 VC 维理论和结构化风险最小原则基础上,具有学习速度快,泛化性好的特点,能够较好地解决小样本、非线性、高维数和局部极小点等问题,已被广泛地运用于识别和分类等实际问题^[12]。SA 是一种基于蒙特卡罗迭代求解法的一种启发式随机搜索算法,该算法有很强的全局优化搜索能力,不要求具备连续性、可导性等假设,算法不仅能够接受优解,也能够接受劣解,这使得 SA 比 GA 较容易跳出局部最优解,从而收敛到全局最优解^[13]。为了尽量防止过拟合现象,使得模型更具有稳健性和更精确的识别率,本文提出一种新方(SA-LS-SVM),该方法采用 LS-SVM 作为识别器,直接采用识别率作为目标函数来寻找特征波数以及所对应的特征波长。为了验证算法的正确性和优越性,分别用 PCA、PLS 以及 SA-LS-SVM 对三种不同品牌的润滑油的可见近红外光谱数据进行处理,用处理完的数据建立模型,然后结合 BP-ANN 对其预测识别,结果发现本文提出的算法的识别率达到 100%,而其他方法都未能达到这个识别率。

2 处理方法和参数设置

2.1 模拟退火算法

SA 是 20 世纪 80 年代初发展起来的一种随机组合优化方法。该算法来源于固体退火原理,将固体加温至充分高,再让其徐徐冷却,在每个温度都达到平衡态,最后在常温时达到基态,内能减为最小。根据 Metropolis 准则,粒子在温度 T 时趋于平衡的概率为 $\exp[-\Delta E/(kT)]$,其中 E 为温度 T 时的内能, ΔE 为其改变量, k 为 Boltzmann 常数。用固体退火模拟组合优化问题,先确定初始温度,随机选择一个初始状态并考察该状态的目标函数值;然后在当前解的领域中,以一定概率选择一个非局部最优解,并令这个解再重复下去,从而不会陷入局部最优,算法有一个控制参数 T 决定,经过大量解变换后,可求得给定控制参数 T 时优化问题的相对最优解。然后缓慢减小控制参数 T 的值,重复上述迭代过程。当 T 逐渐减少并趋于 0 时,最终得到问题的全局最优解^[14]。

2.2 最小二乘向量机(LS-SVM)

LS-SVM 是 SVM 在二次损失函数下的一种改进形式,它把实际问题转化为求解一组线性方程组的问题,简化了计算,提高了收敛速度。LS-SVM 可以这样描述:假定有这样的一组输入集 $\{x_k, y_k\}_{k=1}^N$, $x_k \in \mathbf{R}^d$ 代表输入空间的数据, y_k 代表输出空间的数据。如果输入数据是非线性可分的情况下,可以用一个非线性变换 $\varphi(\cdot): \mathbf{R}^d \rightarrow \mathbf{R}^h$, 将输入的空间数据映射高维的特征空间^[12]。最后在此空间构成最优线性决策函数:

$$y(x) = \text{sgn}[\omega^T \varphi(x) + b], \quad (1)$$

式中 $\omega \in \mathbf{R}^d$ 为超平面权重向量, b 为偏置项。

根据结构风险最小化原则, LS-SVM 最优化模型可描述为

$$\min J(\omega, \xi) = \frac{1}{2} \omega^T \cdot \omega + c \frac{1}{2} \sum_{i=1}^l \xi_i^2, \quad (2)$$

约束条件为

$$y_i = \omega^T \cdot \varphi(x_i) + b + \xi_i, \quad i = 1, \dots, l \quad (3)$$

根据(3)式,将模型转换到对偶空间加以解决,引入 Lagrange 函数:

$$L(\omega, b, \xi, a, \gamma) = \frac{1}{2} \omega^T \cdot \omega + c \frac{1}{2} \sum_{i=1}^l \xi_i^2 - \sum_{i=1}^l a_i [\omega^T \cdot \varphi(x_i) + b + \xi_i - y_i], \quad (4)$$

其中 $a_i \in \mathbf{R}$ 是 Lagrange 乘子, T 表示转置,对(4)式各个变量求偏导,得到以下条件等式:

$$\begin{cases} \partial L / \partial w = 0 \rightarrow w = \sum_{i=1}^l a_i \varphi(x_i), \\ \partial L / \partial \xi_i = 0 \rightarrow a_i = c \xi_i, & i = 1, \dots, l \\ \partial L / \partial b = 0 \rightarrow \sum_{i=1}^l a_i = 0, & i = 1, \dots, l \\ \partial L / \partial a = 0 \rightarrow y_i = w^T \cdot \varphi(x_i) + b + \xi_i. & i = 1, \dots, l \end{cases} \quad (5)$$

从(5)式中消去 w 和 ξ , 可得以下线性方程:

$$\begin{bmatrix} 0 & \mathbf{I}^T \\ \mathbf{I} & \mathbf{\Omega} + \gamma^{-1} \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}, \quad (6)$$

式中 I 为单位矩阵;

$$\begin{aligned} \mathbf{I} &= [1, \dots, 1]; \\ \mathbf{y} &= [y_1, \dots, y_l]; \\ \mathbf{a} &= [a_1, \dots, a_l]; \\ \mathbf{\Omega} &= \{\Omega_{kj} \mid k, j = 1, \dots, l\}. \end{aligned} \quad (7)$$

利用最小二乘法从(6)式解出 \mathbf{a} 和 \mathbf{b} , 可以得到

$$y(x) = \sum_{k=1}^l a_k \varphi^T(x) \varphi(x_k) + b, \quad (8)$$

并且, 定义核函数 $K(x_k, x_j)$:

$$K(x_k, x_j) = \varphi(x_k)^T \varphi(x_j), \quad k, j = 1, \dots, l \quad (9)$$

最后, 就可以得到 LS-SVM 的预测模型:

$$y(x) = \sum_{i=1}^l a_i K(x, x_i) + b. \quad (10)$$

2.3 基于 SA-LS-SVM 识别的特征波长寻优方法

因为 SA 比 GA 更容易找出全局最优点, 而 LS-SVM 支持向量机是一种新型的非线性建模方法, 结合这两种算法的优点, 提出新的一种寻找特征波长的方法, 该方法的主要实现步骤如下:

1) 设置初值: 设置所要搜寻的特征波长个数 N (一般最初设置为 1), 以及识别率初值 P_0 (最初设置为 0);

2) SA 搜寻特征波长, 利用 LS-SVM 对搜寻到的特征波长所对应的光谱数据进行建模, 并对此进行交叉验证得出识别率 P_1 ;

3) 对于 N 个数特征波长, 如果算法满足当前循环的停止条件, 则 $N = N + 1$, 并记下当前特征波长数目的最优识别率 $P(N)$ 以及所对应的特征波长 Spec;

4) 如果识别率达到 100%, 或者 $P(N+1)$ 小于 $P(N)$, 则转到第 5 步, 否则转到第 2 步;

5) 停止算法, 输出最优的特征波长个数 optimNum 和对应的特征波长 optimSpec .

2.4 算法的参数设置

在 SA 中有几个重要参数需要预先设置: 最初

温度, 退火规律, 算法停止条件以及产生新点的方法。本实验中, 设置最初的温度 T 为 100, 退火系数为 0.95^k , 其中 k 为算法的迭代次数, 用 t 分布产生一个扰动点, 对应于固定搜寻特征波长个数算法的停止条件: 在迭代 50 次, 目标函数值变化不超过 0.0001。算法的最终停止条件: 识别率达到 100%, 或者 $N+1$ 个特征波长数目的最优识别率 $P(N+1)$ 小于 N 个波长数目的最优识别率 $P(N)$ 。

而使用 LS-SVM 过程, 需要解决以下两个问题: 选择合适的核函数和最优核参数。现在比较常用的几个核函数有: 径向基核函数 (Radial basis function, RBF) 和多层感知器核函数 (Multi-layer perception, MLP)。在本实验中, 选择的核函数为 RBF 核函数:

$$K(x, x_k) = \exp\{-\|x - x_k\|^2 / \sigma^2\}, \quad (11)$$

选择合适的核函数参数可以提高模型的稳定性和预测率。本实验采用网格搜索法 (Grid-Search) 和留一交叉验证 (Leave Out Cross-Validation) 技术寻找最优参量: 正则化参量 $\text{gam}(\gamma)$ 和 RBF 核函数参量 $\text{sig}2(\sigma^2)$, 这两个参量在很大程度上决定了最小二乘法支持向量机的学习能力和预测能力^[15]。其中 $\text{gam}(\gamma)$ 对保持 LS-SVM 全局执行能力有重要影响, 它主要保持结构风险与最小经验风险平衡; 而 $\text{sig}2(\sigma^2)$ 主要控制 LS-SVM 模型的误差值, 直接

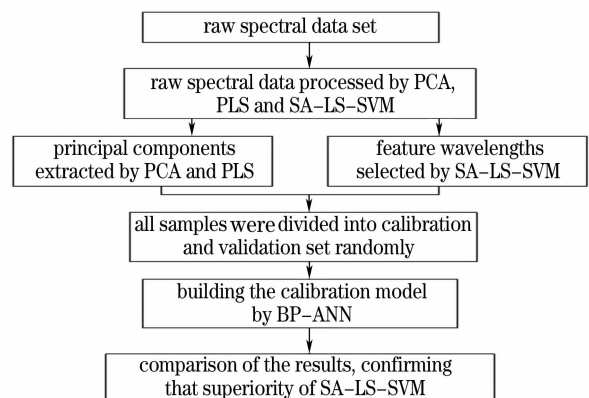


图 1 数据处理的流程图

Fig. 1 Flow chart of data processing

影响最初输入特征值数目^[12]。在本实验中,设置 $\text{gam}(\gamma)$ 的范围为 $2^{-1} - 2^{10}$, 而 $\text{sig}_2(\sigma^2)$ 设置为 $2 - 2^{15}$ ^[16]。SA 和 LS-SVM 模型的运行环境为 Matlab 7.5(The Math Works, Natick, USA)。

2.5 数据处理

本实验的数据处理流程如图 1 所示。

3 实验部分

3.1 样品采集与仪器

实验选用长城、华夏阳光和加德士(下文用种类 1、种类 2 和种类 3 表示)三种品牌的润滑油,采集 180 个样本。盛样容器采用直径为 75 mm, 高为 10 mm 的玻璃皿,样品的采集过程中玻璃皿内润滑油液面高度保持一致,以减少实验误差。光谱仪置于玻璃皿的上方,探头视场角为 20° , 对每一份润滑油样本扫描 30 次。

实验装置由微机、光谱仪、卤素灯、直径 75 mm 的玻璃皿等组成。光谱仪使用美国 ASD (Analytical Spectral Device) 公司的 Handheld Field Spec 光谱仪,光谱采样间隔为 1.5 nm, 采样范围为 325~1075 nm, 扫描次数 30 次,分辨率 3.5 nm,探头视场角为 20° , 采用漫反射方式进行样品光谱采样;光源采用与光谱仪配套的 14.5 V(光源电压)卤素灯。光谱数据以 ASCII 码形式导出并用 ASD View Spec Pro, Unscramble V9.6 (CAMO PROCESS ASOSLO, Norway) 和 DPS (data procession system for practical statistics)软件处理。

3.2 原始光谱数据

三种品牌的润滑油的可见-近红外原始光谱谱线如图 2 所示。图中可以看到,三种润滑油的原始光谱曲线比较相近,不容易区分。图中横坐标为波长,范围为 325~1075 nm,纵坐标为光谱的吸收度。

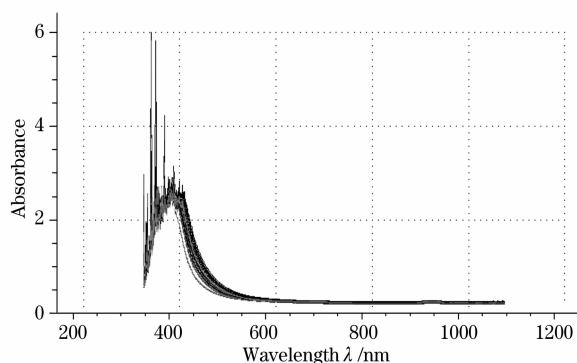


图 2 润滑油的可见/近红外光谱原始图谱

Fig.2 Original Vis/NIR spectroscopy for lubricant

4 结果与讨论

4.1 光谱处理与结果

同时对 180 个样本进行 PCA、PLS 预处理分析,分别用 PCA 和 PLS 提取的前 20 个主成份的特征值及累计可信度如表 1 所示。从表中可以看出,用 PCA 和 PLS 方法得到前 15 个主成分的累计可信度分别为 98.52% 和 98.17%,而后 5 个的成分的累计可信度加起来不超过 1%,因此可以认为前 15 个主成分可以表示原始可见-近红外光谱的主要信息。

表 1 PCA 和 PLS 提取的 20 个主成分的累计可信度

Table 1 First 20 principal components and accumulative reliabilities for PCA and PLS

Principal component	Reliability (PCA)	Principal component	Reliability (PLS)
PC_01	81.96%	PC_01	83.59%
PC_02	85.98%	PC_02	86.74%
PC_03	89.53%	PC_03	86.91%
PC_04	91.67%	PC_04	89.94%
PC_05	93.14%	PC_05	92.14%
PC_06	94.30%	PC_06	92.96%
PC_07	95.39%	PC_07	94.53%
PC_08	96.28%	PC_08	95.19%
PC_09	96.88%	PC_09	96.12%
PC_10	97.27%	PC_10	96.95%
PC_11	97.62%	PC_11	97.35%
PC_12	97.92%	PC_12	97.59%
PC_13	98.17%	PC_13	97.75%
PC_14	98.36%	PC_14	98.05%
PC_15	98.52%	PC_15	98.17%
PC_16	98.66%	PC_16	98.24%
PC_17	98.78%	PC_17	98.26%
PC_18	98.89%	PC_18	98.19%
PC_19	98.99%	PC_19	98.21%
PC_20	99.06%	PC_20	98.12%

在 SA-LS-SVM 中,首先从 180 个样本随机选择 90 个,然后用 SA 对其进行特征波长的搜寻。选用识别率作为目标函数,如果识别率达到 100%,认为是最佳预测值,特征波长寻优结果如图 3 所示,图 3(a)为寻找的最优光谱数据点的序号,纵坐标 0 对应于真实光谱的起始波长 325 nm,寻优范围为 0~751(对应于真实波长范围 325~1075 nm)。图 3(b)是迭代次数,以及每次迭代所得到的目标函数值。图 3(c)为当前寻到的波长,纵坐标的函数值为识别率值,由于此设计的退火算法程序是寻最小值,而要寻的识别率为最大值,所以纵坐标值(目标函数值)取负值。图 3(d)为 LS-SVM 建模时, σ^2 和 γ 的最优值搜寻图。从图 3 中可以看出,基于 SA-LS-SVM,

找到 4 个波长就可以把这三种品牌的润滑油完全区分开。所找的波长分别为 80 nm、159 nm、569 nm 和 675 nm(真实波长为 405 nm、484 nm、894 nm 和 1000 nm)。

对于 PCA 和 PLS 预处理后的光谱数据,用得到 15 个主要成分代替原始光谱的全部信息,而对于 SA-LS-SVM,用提取 4 个最优波长对于的光谱数据代替原始光谱信息。预处理后的 180 个样本,随机

分成建模集和预测集各 90 个(每种品牌 30 个),然后用 BP-ANN 进行建模和预测,识别结果见表 2。BP-ANN 在 DPS V3.11 专业版数据处理系统环境下运行,基本参数设置采用原系统的默认值,隐含网络层数 3,输入层节点数 10,最小建模速率 0.1,动态参数 0.6,参数 Sigmoid 0.9,允许误差 0.0001,最大迭代次数 1000,数据转换方法为标准化变换。

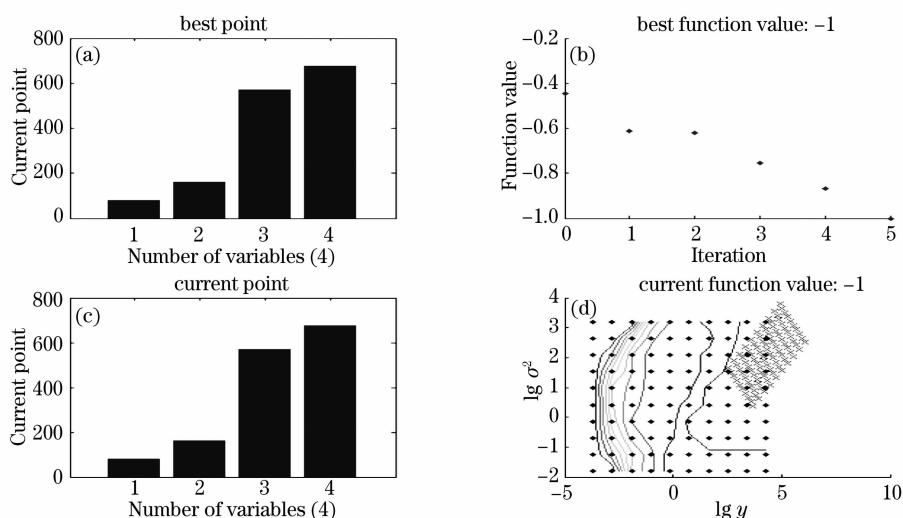


图 3 用 SA 寻找特征波长结果图

Fig. 3 Results of characteristic wavelengths by SA-LS-SVM

表 2 不同方法结合 BP-ANN 的识别结果

Table 2 Discrimination results with combination of different methods with BP-ANN

	Variety	Number	Variety			Discriminating rate of each variety	Total discriminating rate
			1	2	3		
PCA & BP-ANN	1	30	27	1	2	90.00%	72.22%
	2	30	0	11	19	36.67%	
	3	30	1	2	27	90.00%	
PLS & BP-ANN	1	30	30	0	0	100.00%	90.00%
	2	30	0	21	9	70.00%	
	3	30	0	0	30	100.00%	
SAA-LS-SVM & BP-ANN	1	30	30	0	0	100.00%	100.00%
	2	30	0	30	0	100.00%	
	3	30	0	0	30	100.00%	

4.2 结果讨论

本实验运用不同的处理方法对预测集进行识别,取得不同的结果,结果可以从表 2 中可以得到。从表 2 中可以看出,SA-LS-SVM & BP-ANN 预测率达到 100%,PLS & BP-ANN 的预测率也不错,达到 90%。预测率最差的是 PCA & BP-ANN,预测率仅仅为 72.22%。进一步,也可以从表 2 中看出,不管用哪种算法(除了 SA-LS-SVM 外)对三种润滑油进行预测,第二种品牌的润滑油的预测率都

是最低,而第一种和第三种品牌的润滑油的预测率都比较高。在 PCA & BP-ANN 模型中,把第一种品牌的润滑油误预测成第二种品牌有 1 个,误预测成第三种品牌有 2 个,把第二种品牌的润滑油误预测成第三种品牌有 19 个,把第三种品牌的润滑油误识别成第二种品牌的润滑油有 2 个,误预测成第一种品牌有 1 个。在 PLS & BP-ANN 模型中,只有在第二种的预测出现了错误,把第二种品牌的润滑油误预测成第三种有 9 个。总之,从表中可以得出

这样的结果,在本次实验中,SA-LS-SVM & BP-ANN 预测效果最好,其次是 PLS & BP-ANN 模型,效果最差的是 PCA & BP-ANN 模型,在 PCA & BP-ANN 和 PLS & BP-ANN,第一种和第三种品牌的润滑油的预测效果比较理想,而第二种品牌的润滑油的预测效果比较差。

5 结 论

LS-SVM 是一种新型的非线性建模方法,具有学习速度快、泛化性好的特点,用它作为 SA 寻找最优波长的预测器,得到的识别率作为目标函数,不仅能有效的减少建模的变量数,而且可以提高预测精度。这两种方法的结合使用,突破了以往选择特征谱区的主观性,使特征谱区的选择更科学化。为了验证本文提出的算法优越性,把 SA-LS-SVM、PCA 和 PLS 提取的特征波长或主成分结合 BP-ANN,对 3 种品牌的润滑油进行了预测。从实验结果来看,只有 SA-LS-SVM & BP-ANN 的预测率达到了 100%,而其他任何一种模型都没有达到 100%。说明针对于光谱技术品种识别问题,本文提出的 SA-LS-SVM 比其他几个传统方法提高了识别率。所以,该结果可为近红外快速检测仪提供一种客观选择特征波长的方法。

参 考 文 献

- 1 Shao Yongni, He Yong, Feng Shuijuan. Measurement of yogurt internal quality through using Vis/NIR spectroscopy[J]. *Food Res. Int.*, 2007, **40**(7): 835~841
- 2 Chen Quansheng, Zhao Jiewen, Zhang Haidong *et al.*. Identification of authenticity of tea with near infrared spectroscopy based on support vector machine[J]. *Acta Optica Sinica*, 2006, **26**(6): 933~937
陈全胜,赵杰文,张海东等. 基于支持向量机的近红外光谱鉴别茶叶的真伪[J]. *光学学报*, 2006, **26**(6): 933~937
- 3 Shao Yongni, He Yong, Wang Yanyan. A new approach to discriminate varieties of tobacco using vis/near infrared spectra [J]. *Eur. Food Res. Technol.*, 2007, **224**(5): 591~596
- 4 He Yong, Feng Shuijuan, Deng Xunfei *et al.*. Study on lossless discrimination of varieties of yogurt using the Visible/NIR-spectroscopy[J]. *Food Res. Int.*, 2006, **39**(6): 645~650
- 5 Zhu Shiping, Wang Yiming, Zhang Xiaochao *et al.*. Region selection method of near infrared spectrum based on genetic algorithm[J]. *Transactions of CSAM*, 2004, **35**(5): 152~156
祝诗平,王一鸣,张小超等. 基于遗传算法的近红外光谱谱区选择方法[J]. *农业机械学报*, 2004, **35**(5): 152~156
- 6 Li Xiaoli, He Yong, Fang Hui. Non-destructive discrimination of Chinese bayberry varieties using Vis/NIR spectroscopy[J]. *J. Food Engng.*, 2007, **81**(2): 357~363
- 7 Cen Haiyan, Bao Yidan, He Yong. Fast discrimination of varieties of bayberry juice based on spectroscopy technology[J]. *Spectrosc. Spectr. Anal.*, 2007, **27**(3): 503~506
岑海燕,鲍一丹,何勇. 基于光谱技术的杨梅汁品种快速鉴别方法的研究[J]. *光谱学与光谱分析*, 2007, **27**(3): 503~506
- 8 Zou Xiaobo, Zhao Jiewen. Methods of characteristic wavelength region and wavelength selection based on genetic algorithm[J]. *Acta Optica Sinica*, 2007, **27**(7): 1316~1321
邹小波,赵杰文. 用遗传算法快速提取近红外光谱特征区和特征波长[J]. *光学学报*, 2007, **27**(7): 1316~1321
- 9 D. Jouan-Rimbaud, D. L. Massat, O. E. de Noord. Random correlation in variable selection for multivariate calibration with a genetic algorithm[J]. *Intell. Lab. Syst.*, 1996, **35**(2): 213~220
- 10 U. Depczynski, K. Jetter, K. Molt *et al.*. Quantitative analysis of near infrared spectra by wavelet coefficient regression using a genetic algorithm[J]. *Intell. Lab. Syst.*, 1999, **47**(2): 179~187
- 11 Lan Hai, Wang Xiong, Wang Ling. Improved genetic-annealing algorithm for global optimization of complex functions[J]. *J. Tsinghua University (Sci & Tech)*, 2002, **42**(9): 1237~1240
蓝海,王雄,王凌. 复杂函数全局最优化的改进遗传退火算法[J]. *清华大学学报(自然科学版)*, 2002, **42**(9): 1237~1240
- 12 Wu Di, He Yong, Feng Shuijuan *et al.*. Study on infrared spectroscopy technique for fast measurement of protein content in mild powder based on LS-SVM[J]. *J. Food Engng.*, 2008, **84**(1): 124~131
- 13 Z. G. Wang, Y. S. Wong, M. Rahman. Development of a parallel optimization method based on genetic simulated annealing algorithm[J]. *Parallel Comput.*, 2005, **31**(8~9): 839~857
- 14 Richard O. Duda, Peter E. Hart, David G. Stork. *Pattern Classification* [M]. Citic Publishing House, second edition, Beijing; China Machine Press, 2003. 284~291
理查德,皮特,大卫. 模式预测[M]. 第二版,李宏东,姚天翔等译,北京:机械工业出版社,中信出版社,2003. 286~291
- 15 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**(10): 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**(10): 2054~2058
- 16 Zhu Jiayuan, Yang Yun, Zhang Hengxi *et al.*. Data prediction with few observations based on optimized least squares support vector machines[J]. *Acta Aeronautica et Astronautica Sinica*, 2004, **25**(6): 565~568
朱家元,杨云,张恒喜等. 基于优化最小二乘支持向量机的小样本预测研究[J]. *航空学报*, 2004, **25**(6): 565~568