

近红外光谱结合极限学习机和 GA-PLS 算法 检测普洱茶茶多酚含量

张海东 李贵荣* 李若诚 许文方 华英杰

(云南农业大学工程技术学院, 云南 昆明 650201)

摘要 利用近红外光谱技术检测普洱茶中茶多酚的含量,首先通过遗传偏最小二乘法(GA-PLS)筛选出表征茶多酚含量的特征波数点,并进行主成分分析,然后建立极限学习机(ELM)预测模型。研究得到的最佳 ELM 预测模型涉及 40 个变量,主成分分析后以第 1、第 2 主成分作为输入,以 Sigmoidal 函数为隐含层神经元激励函数,隐含层神经元个数确定为 13。模型的交互验证均方根误差值、预测集均方根误差值和预测集相关系数 R^2 分别为 1.0109、1.6686 和 0.9705,预测性能明显优于全光谱偏最小二乘模型和遗传偏最小二乘模型。说明利用近红外光谱技术结合极限学习机和遗传偏最小二乘法可以很好地预测普洱茶中茶多酚的含量。

关键词 光谱学;近红外光谱;极限学习机;遗传偏最小二乘法;普洱茶;茶多酚含量

中图分类号 O657.33 文献标识码 A doi: 10.3788/LOP50.043001

Determination of Tea Polyphenols Content in Puerh Tea Using Near-Infrared Spectroscopy Combined with Extreme Learning Machine and GA-PLS Algorithm

Zhang Haidong Li Guirong Li Ruocheng Xu Wenfang Hua Yingjie

(Faculty of Engineering and Technology, Yunnan Agricultural University, Kunming, Yunnan 650201, China)

Abstract To determine the tea polyphenols content in Puerh tea by near infrared spectroscopy, genetic algorithm combined with partial least square (GA-PLS) is used to select those wave-numbers carrying information that are highly related to the tea polyphenols of Puerh tea. As a result, 40 wave numbers are selected from the spectral range of $10001\sim 4000\text{ cm}^{-1}$. Then principal component analysis is applied for these selected wave numbers, and the first two principal components are achieved to input into the extreme learning machine (ELM). Using sigmoidal function as active function, the ELM is trained with the number of hidden-layer neurons varying from 1 to 50. Result shows that it will reach the minimum root mean square error of prediction (RMSEP) set and get the optimal ELM model when there are 13 hidden-layer neurons. The optimal ELM model gives the correlation coefficient of prediction set R^2 of 0.9705, with root mean square error of cross-validation (RMSECV) of 1.0109 and RMSEP of 1.6686. This model uses only 40 wave numbers and 2 input-layer neurons, but obtains better performance, compared with the global PLS model with 1557 wave numbers and GA-PLS model with 40 wave numbers. Results show that the tea polyphenols content in Puerh tea can be determined with high precision by using near-infrared spectroscopy combined with genetic algorithm and ELM.

Key words spectroscopy; near infrared spectroscopy; extreme learning machine (ELM); genetic algorithm-partial least square algorithm (GA-PLS); Puerh tea; tea polyphenols content

OCIS codes 300.6340;200.4260;120.4290

1 引言

普洱茶是中国云南特有的茶品,由大叶种茶叶经特殊发酵工艺加工而成,具有独特的滋味和香气。研究

收稿日期:2012-10-31;收到修改稿日期:2013-01-24;网络出版日期:2013-03-14

基金项目:国家自然科学基金(30760103)资助课题。

作者简介:张海东(1974—),男,博士,副教授,主要从事农产品品质无损检测方面的研究。E-mail: zhd_74@126.com

*通信联系人。E-mail: ynlgr@yahoo.com.cn

表明,普洱茶特有的滋味和香气是在后发酵过程中,以茶多酚为主的多种化学成分在微生物和酶的作用下形成的。因此,茶多酚含量是评价普洱茶品质的一项重要指标。茶多酚是从茶叶中提取的由多种儿茶素单体及少量咖啡碱以及其他物质组成的混合物。传统上,茶多酚含量主要依靠化学分析法检测,但该法操作繁琐、费时,且对分析人员的要求较高。近些年来,近红外光谱技术被越来越多地应用于包括茶叶在内的农产品的内部品质检测中^[1~5],并都取得了很好的效果。在茶叶中茶多酚含量的近红外光谱检测方面,孙耀国等^[4]用近红外光谱法检测绿茶中氨基酸、咖啡碱和茶多酚的含量,通过将光谱分成三个波段并分析各波段偏最小二乘(PLS)预测模型后发现 5000~4000 cm^{-1} 波段的光谱结合一阶导数预处理法对茶多酚含量的 PLS 预测效果最佳。罗一帆等^[5]利用近红外光谱法测定茶粉中的茶多酚和茶多糖,根据光谱特点选择 7432.3~6155.7 cm^{-1} 和 5484.6~4192.5 cm^{-1} 两个波段共 668 个波数点建立人工神经网络(ANN)预测模型。Chen 等^[6]利用近红外光谱法同时测定绿茶中总茶多酚和咖啡碱含量,并分析了标准正态变量变换(SNV)、一阶导数、二阶导数法分别预处理光谱后,PLS 预测模型的性能。刘辉军等^[7]通过分析光谱特点,选择 1400~1628 nm 和 1679~2000 nm 两个波段建立茶叶中茶多酚含量的近红外光谱 PLS 预测模型。罗文文等^[8]研究了绿茶中水分和茶多酚含量的近红外光谱定标模型,分析比较了主成分回归(PCR)模型和 PLS 模型的性能以及不同预处理方法对模型性能的影响。Chen 等^[9]利用区间偏最小二乘法(iPLS)和联合区间偏最小二乘法(siPLS)分别筛选绿茶中茶多酚含量的近红外建模区间,并与全光谱偏最小二乘模型进行比较,发现利用 siPLS 法筛选出的区间所建 PLS 模型性能最佳。林新等^[10]研究了绿茶中茶多酚含量的近红外光谱预测模型,利用正交试验法分析比较了多元散射校正(MSC)等光谱预处理方法对 PCR、PLS 以及改进的偏最小二乘(MPLS)模型性能的影响。

可以看出,学者们的研究主要侧重两个方面:一是建模波长的选择,二是对建模方法的分析比较。对于建模的波长,研究者或根据光谱特点(文献[5,7]),或结合建模方法(文献[4,9])获得了用于预测茶叶中茶多酚含量的光谱区间,但所筛选出的区间或联合区间仍不可避免地会将许多与茶多酚含量无关的信息以及冗余信息带入模型,且容易将落选区间中与被测组分高度相关的光谱信息排除在外,影响模型的预测精度和分析效率。对于建模方法,学者们采用最多的是偏最小二乘法,但对于茶叶这种含有多种组分的非线性复杂体系,偏最小二乘法并不是预测其内部品质的最佳算法。而常用的人工神经网络法虽然能获得很好的预测模型^[5],但在网络训练过程中需要不断演算、调整各层之间的连接权值、隐含层神经元数目以及隐含层神经元的阈值以获得最佳的网络参数,因此模型训练速度较慢。

本研究将近红外光谱技术用于普洱茶中茶多酚含量的检测,利用遗传偏最小二乘法(GA-PLS)筛选建模波长,然后输入新型神经网络——极限学习机(ELM)网络进行训练,以期获得更为简单、适用的茶多酚含量预测模型,克服现有相关研究中存在的不足。

2 算法简介

2.1 极限学习机理论

极限学习机是 Huang 等^[11]由单隐含层前馈神经网络(SLFN)发展起来的一种新型神经网络算法,目前已在回归预测^[12~15]、模式识别^[16~18]等领域取得了很好的应用效果。该算法解决了 SLFN 网络训练速度慢、容易陷入局部极小点以及对学习率的选择敏感等缺点。极限学习机算法理论简述如下。

Huang 等^[11]发现,在 SLFN 中,当隐含层神经元的激励函数 $g(x)$ 无限可微时,若隐含层神经元个数 K 与训练集样本数 N 相等,则对任意的输入层神经元与隐含层神经元的连接权值 w 和隐含层神经元的偏置 b ,SLFN 网络都可以零误差逼近训练样本。或者当隐含层神经元个数 K 小于训练集样本数 N 时,对任意的 w 和 b ,SLFN 网络的训练误差可以逼近一个任意小的数 ϵ ($\epsilon > 0$)。在此条件下, w 和 b 只需在 SLFN 网络训练前随机选取,在训练过程中则可保持不变。而隐含层与输出层之间的连接权值 β 可通过最小二乘法求解以下问题获得:

$$\min_{\beta} \| H\beta - T \|, \quad (1)$$

式中 H 为隐含层输出矩阵, T 为 SLFN 网络的输出矩阵。

可解得

$$\hat{\beta} = \mathbf{H}^+ \mathbf{T}, \quad (2)$$

式中 \mathbf{H}^+ 为隐含层输出矩阵 \mathbf{H} 的 Moore-Penrose 广义逆。

由上述分析可知,对于给定的样本集 $\{\mathbf{x}_i, \mathbf{y}_i\}, i=1, 2, \dots, N$, 极限学习机的算法步骤为:1)确定隐含层神经元个数,选择隐含层神经元的激励函数,并随机设定输入层与隐含层间的连接权值 $w_j, j=1, 2, \dots, l$ 和隐含层神经元的偏置 $b_j, j=1, 2, \dots, l$;2)计算隐含层输出矩阵 \mathbf{H} ;3)计算输出层权值 $\hat{\beta} = \mathbf{H}^+ \mathbf{T}$ 。

2.2 遗传偏最小二乘法

遗传偏最小二乘法由 Leardi^[19,20] 提出,该算法能在全光谱区范围内筛选出表征待测品质信息的特征波数。算法的基本思想是把全光谱区的每一个波数点作为一个基因进行 0/1 二进制编码(1 表示波数点被选中,0 表示未被选中),将所有基因的二进制编码排列在一起就构成了一条染色体。然后通过选择、交叉和变异等算子进行遗传迭代,迭代过程中,用适应度函数评价每次迭代被选中的波数点所建偏最小二乘模型的性能。迭代结束后将被选用频数最高的前 n (n 为预先设定)个波数点按选用频数高低逐一加入待测品质的偏最小二乘预测模型,以最小的交互验证均方根误差(RMSECV)值确定出最佳的建模波数。遗传迭代过程中的适应度函数可以用 RMSECV 值^[21]、预测相关系数^[22]或预测相对误差^[23]为评价目标来构建。

3 实验方法与数据

3.1 实验材料

实验共采集云南勐海、绿春、下关、思茅、临沧等地不同厂家生产的普洱茶样本 70 个,其中生茶样本 34 个,熟茶样本 36 个。从中随机抽取 50 个作为校正集,余下的 20 个作为预测集,分别编号后避光贮存。

3.2 光谱采集

样本近红外光谱利用 Antaris II 型傅里叶变换近红外光谱仪(美国 Thermo Fisher 公司)采集,采集方法为漫反射法。采集时,光谱仪参数设置为:光谱范围 $10001 \sim 4000 \text{ cm}^{-1}$,扫描次数 16 次,分辨率 8 cm^{-1} ,数据采样间隔 3.856 cm^{-1} 。每条光谱含有 1557 个波数点。实验过程中,室内温度控制在 $26 \text{ }^\circ\text{C}$ 。每个样本分别采集 3 次光谱,取 3 次的平均值作为原始光谱。光谱采集前不对普洱茶样本进行粉碎和过筛等预处理。考虑到光谱采集时系统、环境以及样本颗粒大小不均等因素对近红外光谱的影响,在波数筛选和建模前需要对原始光谱进行预处理,以减小上述因素的影响。本研究经过分析比较,对原始光谱进行中心化预处理。图 1 为校正集原始光谱和中心化后的光谱。

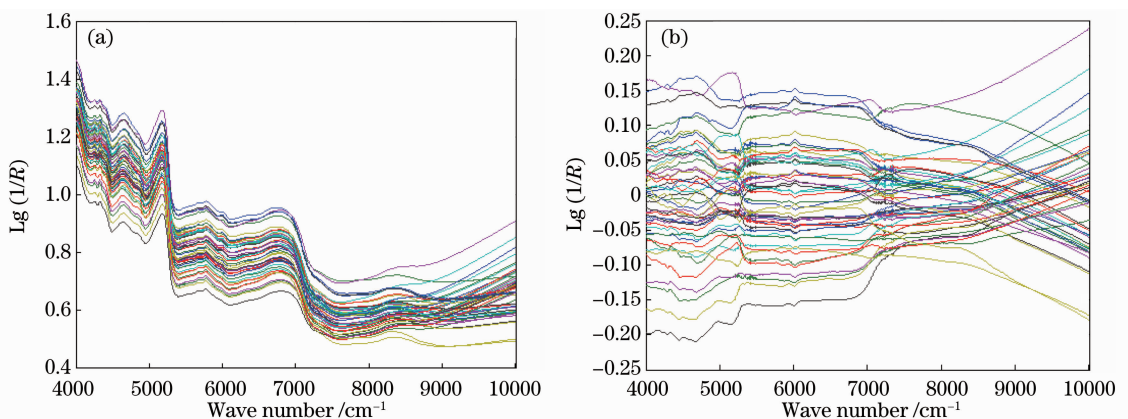


图 1 (a)校正集原始光谱和(b)中心化处理后的光谱

Fig. 1 (a) Raw spectra and (b) mean centered spectra of calibration set

3.3 茶多酚含量化学检测

光谱采集后,茶多酚含量的化学分析按国家标准 GB/T 8313-2002 中规定的方法执行,即利用茶叶中多酚类物质与亚铁离子形成紫蓝色络合物的原理进行分光光度检测。表 1 分别列出了校正集与预测集样本中茶多酚含量实测值的范围、均值和标准偏差。

表1 校正集和预测集样本中茶多酚含量实测值

Table 1 Reference measurement of tea polyphenols content in training set and prediction set

Sample set	Number of samples	Range /%	Mean /%	Standard deviation /%
Calibration set	50	3.3~33.9	16.502	10.3309
Prediction set	20	5.4~33.4	19.650	9.9715

4 数据处理与结果分析

光谱中心化处理后先用遗传偏最小二乘法筛选出表征茶多酚含量的特征波数,然后以特征波数为输入训练极限学习神经网络模型。研究中的所有数据处理均在 Matlab 6.5(美国 Mathworks)平台上完成。

4.1 遗传偏最小二乘法筛选特征波数

本研究中,遗传偏最小二乘法的搜索范围为全谱区,即 $10001\sim 4000\text{ cm}^{-1}$,共 1557 个波数点。迭代过程中以 RMSECV 值为评价指标,RMSECV 值越小表明被选波数点所建模型性能越好,由此可构建适应度函数:

$$F(X_k) = \frac{1}{1 + E_{CV}(X_k)}, \quad (3)$$

式中 X_k 代表种群中第 k 个染色体, $E_{CV}(X_k)$ 为第 k 个染色体产生的 RMSECV 值, $F(X_k)$ 为第 k 个染色体的适应度值。遗传算法的控制参数设定为:初始群体大小为 30,交叉概率 $p_c=0.5$,变异概率 $p_m=0.01$,遗传迭代次数为 100 次。迭代结束后将选用频数最高的前 100 个波数点依选用频数高低逐一顺序加入茶多酚 PLS 预测模型,以最小的 RMSECV 值确定建模波数点。为防止遗传操作过程中的随机性对波数筛选结果的影响,共进行了 5 次运算,最后挑选出其中综合性能最好的模型所选用的波数点作为最佳波数点。

表 2 列出了各次运算的波数筛选情况及所建偏最小二乘回归模型的性能。可以看出,5 次运算分别筛选出 48、65、40、66 和 70 个波数,都远远少于全谱区 1557 个波数。从模型的复杂度来看,全谱 PLS 模型(Global PLS model)采用了 6 个主成分,5 个 GA-PLS 模型采用的主成分数分别为 4 个(GA-PLS3)、5 个(GA-PLS2、4)及 7 个(GA-PLS1、5),与全谱模型相比,参与建模的主成分数没有明显变化。从模型的性能看,全谱模型的预测均方根误差(RMSEP)为 2.3746,而 5 个 GA-PLS 模型的 RMSEP 值都比全谱模型提高了 7.8%以上,说明经 GA-PLS 法筛选建模波数后,模型对外部样本的预测能力有了明显的提高。这主要是由于遗传算法高效的全局搜索能力和偏最小二乘法本身所具备的信息提取能力有机结合后,将谱区中与茶多酚含量不相关或相关性极小的波数点剔除的缘故。

表2 遗传偏最小二乘法筛选波数结果

Table 2 Results of wave-number selection calculated from GA-PLS

Model	Number of selected wave numbers	Number of principal components	RMSECV	RMSEP	R^2 for prediction set
Global PLS	1557	6	1.2432	2.3746	0.9433
GA-PLS1	48	7	1.0790	2.1770	0.9523
GA-PLS2	65	5	1.1203	2.1384	0.9540
GA-PLS3	40	4	1.3766	1.7490	0.9692
GA-PLS4	66	5	1.1087	2.1880	0.9518
GA-PLS5	70	7	1.1742	2.0232	0.9588

综合分析 5 次运算后所建 PLS 模型的性能,本研究选取第 3 次运算筛选出的波数作为极限学习神经网络的输入。第 3 次运算共筛选出 40 个波数,图 2 显示了这 40 个波数在谱区中的位置。

4.2 极限学习机预测模型的建立

考虑到校正集样本数为 50,与 GA-PLS 法筛选出的变量数较为接近,为了防止极限学习神经网络产生过拟合,对筛选出的 40 个变量进行主成分分析,以第 1、第 2 主成分(方差贡献率分别为 82.92%和 15.65%)作为 ELM 的输入。根据理论,极限学习神经网络的隐含层神经元数目需在训练前确定,而其输入层与隐含层之间的权值 w 以及隐含层神经元的偏置 b 只需随机设定,且训练过程中无需进行调整。因此网络训练前确定隐含层神经元的数目,本研究中将隐含层神经元数目初始化为 1,并以 1 为步长逐步增加到 50(校正集样

本数),用 RMSEP 值来确定最佳的隐含层神经元数目。本研究以 Sigmoidal 函数为 ELM 网络隐含层神经元的激励函数。图 3 显示了隐含层神经元数目对极限学习机性能的影响。从图中可以看出,随着隐含层神经元数的增加, RMSECV 值逐渐减小,当隐含层神经元数增加到 50,即与校正集样本数相等时, RMSECV 值趋近于 0(为 1.9059×10^{-8})。但从 RMSEP 值来看,随着隐含层神经元数的增加, RMSEP 值呈先减小后增大的变化趋势,当隐含层神经元数为 13 时, RMSEP 达到最小值 1.6686, 隐层神经元数为 50 时, RMSEP 达到最大值 13.7526。根据上述分析,本研究取隐含层神经元数为 13。图 4 为隐含层神经元数取 13 时,利用第 1、第 2 主成分作为输入建立的极限学习机模型(GA-ELM 模型)对校正集和预测集样本的预测值与实测值的散点图,从图中可以看出,该模型对普洱茶中茶多酚的含量具有很高的预测精度。

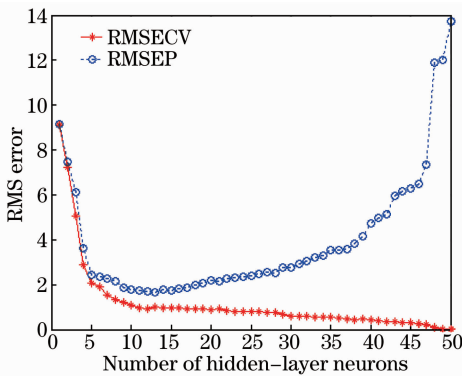


图 3 隐含层神经元个数对极限学习机模型均方根误差的影响

Fig. 3 Effect of hidden-layer neurons number on root mean square error of ELM model

4.3 极限学习机预测模型性能分析

表 3 列出了经波数筛选后的极限学习机预测模型(GA-ELM 模型)和全谱极限学习机模型(Global-ELM)对普洱茶中茶多酚含量的预测性能。从表中可以看出,两种模型都能很好地预测普洱茶中茶多酚的含量,但 GA-ELM 模型的输入层神经元数仅为 2,其 RMSECV 值为 1.0109, RMSEP 值为 1.6686, 预测集相关系数 R^2 为 0.9705, 综合模型的结构、均方根误差以及预测集相关系数等几项指标来看, GA-ELM 模型对普洱茶中茶多酚含量的预测性能明显优于 Global-ELM 模型,其对外部样本的预测能力比后者高出 9.9%,同时大大减少了数据处理量,提高了处理效率。模型也优于全光谱 PLS 模型和 GA-PLS 模型(表 2)。图 5 显示了全谱 PLS 模型、全谱 ELM 模型和 GA-ELM 模型对预测集样本的预测性能,从图中也可以看出, GA-ELM 模型的预测精度最高,全谱 ELM 模型次之,全谱 PLS 模型精度最低。从模型性能参数看, GA-ELM 模型的 RMSECV 值和

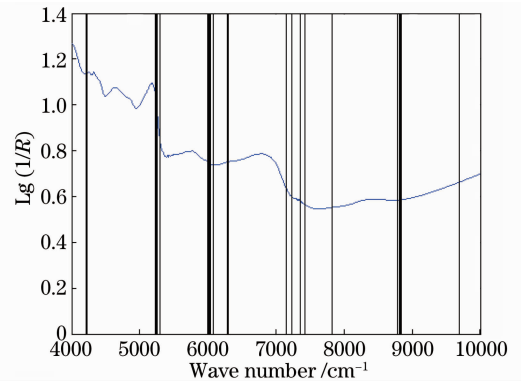


图 2 GA-PLS 法筛选出的最佳波数点在谱区中的位置
Fig. 2 Distribution diagram of GA-PLS selected wave numbers

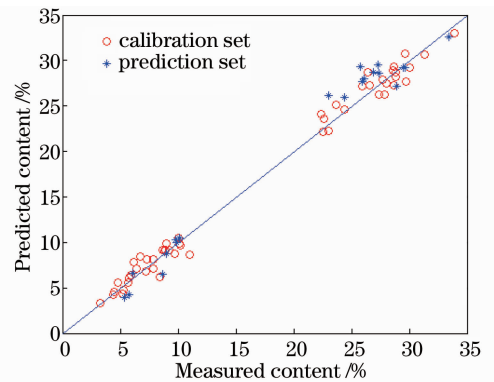


图 4 GA-ELM 模型对校正集和预测集样本的茶多酚预测值与实测值之散点图

Fig. 4 Reference measurement versus GA-ELM prediction of tea polyphenols content for calibration set and prediction set

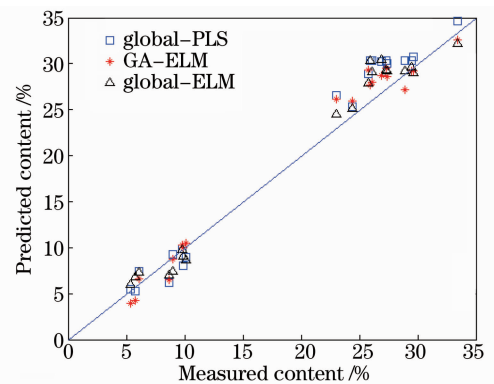


图 5 全谱 PLS 模型、全谱 ELM 模型及 GA-ELM 模型的预测性能比较

Fig. 5 Comparison of prediction precision among global PLS model, global ELM model and GA-ELM model

RMSEP 值分别比全光谱 PLS 模型减小 18.69% 和 29.73%。从对外部样本的预测能力来看,全谱 ELM 模型的性能虽然低于 GA-ELM 模型和表 2 中的 GA-PLS 模型,但仍表现出了较好的预测性能。这说明无论是否经过波数筛选,极限学习神经网络模型都可以很好地预测普洱茶中茶多酚的含量。同时,模型在训练过程中只需对一个参数即隐含层神经元个数进行人为调整,克服了传统网络需要反复调整多个参数带来的不便。

表 3 GA-ELM 预测模型和全谱 ELM 预测模型性能比较

Table 3 Performance comparison of GA-ELM model and global ELM model

Model	Number of selected wave numbers	Number of input-layer neurons	Number of hidden-layer neurons	RMSECV	RMSEP	R ² for prediction set
Global-ELM	1557	1557	12	1.1848	1.8518	0.9637
GA-ELM	40	2	13	1.0109	1.6686	0.9705

5 结 论

利用近红外光谱技术结合极限学习神经网络预测普洱茶中茶多酚的含量,取得了满意的预测精度。特别是经遗传算法筛选波数后建立的茶多酚 ELM 预测模型的各项性能参数都明显优于全谱 ELM 模型、GA-PLS 模型以及传统的全谱 PLS 模型。而且该模型仅涉及 40 个波数点,输入层神经元数为 2,远少于全谱模型的 1557 个波数点,大大减少了数据运算量,提高了分析效率,显示了近红外光谱技术结合极限学习机和遗传偏最小二乘法快速检测普洱茶中茶多酚含量的潜力。

参 考 文 献

- Guo Fei, Liu Jingjing, Luo Xiao *et al.*. Feasibility study of RBF fuzzy neural network in cashmere and wool identification based on near infrared spectroscopy[J]. *Laser & Optoelectronics Progress*, 2012, **49**(8): 083001
郭飞,刘净净,罗宵等. RBF 模糊神经网络用于 NIR 鉴别羊绒和羊毛的可行性研究[J]. *激光与光电子学进展*, 2012, **49**(8): 083001
- Zhao Jiewen, Zhang Yanhua, Chen Quansheng *et al.*. Measurement of TVB-N content by multi-information fusion technique based on spectroscopy and imaging[J]. *Laser & Optoelectronics Progress*, 2012, **49**(6): 063003
赵杰文,张燕华,陈全胜等. 光谱和成像融合技术检测猪肉中挥发性盐基氮[J]. *激光与光电子学进展*, 2012, **49**(6): 063003
- Guo Weiliang, Wang Dan, Song Jia *et al.*. Simultaneous and rapid quantitative analysis of four components in cordyceps militaris mycelium powder using near infrared spectroscopy[J]. *Acta Optica Sinica*, 2011, **31**(2): 0230002
郭伟良,王丹,宋佳等. 近红外光谱法同时快速定量分析蛹虫草菌丝体中 4 种有效成分[J]. *光学学报*, 2011, **31**(2): 0230002
- Sun Yaoguo, Lin Min, Lü Jin *et al.*. Determination of the contents of amino acids, caffeine, and polyphenols in green tea by near infrared spectroscopy[J]. *Chinese J. Spectroscopy Laboratory*, 2004, **21**(5): 939~943
孙耀国,林敏,吕进等. 近红外光谱法测定绿茶中氨基酸、咖啡碱和茶多酚的含量[J]. *光谱实验室*, 2004, **21**(5): 939~943
- Luo Yifan, Guo Zhenfei, Zhu Zhenyu *et al.*. Studies on ANN models of determination of tea polyphenol and amylase in tea by near-infrared spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2005, **25**(8): 1230~1233
罗一帆,郭振飞,朱振宇等. 近红外光谱测定茶叶中茶多酚和茶多糖的人工神经网络模型研究[J]. *光谱学与光谱分析*, 2005, **25**(8): 1230~1233
- Chen Quansheng, Zhao Jiewen, Huang Xinyi *et al.*. Simultaneous determination of total polyphenols and caffeine contents of green tea by near-infrared reflectance spectroscopy[J]. *Microchemical Journal*, 2006, **83**(1): 42~47
- Liu Huijun, Lü Jin, Zhang Weigang *et al.*. A testing model for tea polyphenol content based on NIR[J]. *Infrared Technology*, 2007, **29**(7): 429~432
刘辉军,吕进,张维刚等. 茶叶中茶多酚含量的近红外光谱检测模型研究[J]. *红外技术*, 2007, **29**(7): 429~432
- Luo Wenwen, Zhang Yueling, Gong Shuying *et al.*. Preliminary construction of NIR calibration model for moisture and total tea polyphenols of green tea[J]. *Journal of Tea*, 2007, **33**(2): 67~70
罗文文,张月玲,龚淑英等. 绿茶水分和茶多酚总量近红外分析定标模型的建立与应用[J]. *茶叶*, 2007, **33**(2): 67~70
- Chen Quansheng, Zhao Jiewen, Liu Muhua *et al.*. Determination of total polyphenols content in green tea using FT-NIR spectroscopy and different PLS algorithms[J]. *J. Pharmaceutical and Biomedical Analysis*, 2008, **46**(3): 568~573
- Lin Xin, Niu Zhiyou. Optimization of models for quantitative analysis of tea polyphenols in green tea by near infrared spectrophotometry[J]. *Food Science*, 2009, **30**(10): 144~148
林新,牛智有. 绿茶茶多酚近红外光谱定量分析模型优化研究[J]. *食品科学*, 2009, **30**(10): 144~148

- 11 G. B. Huang, Q. Y. Zhu, C. K. Siew. Extreme learning machine; theory and applications[J]. *Neurocomputing*, 2006, **70**(1-3): 489~501
- 12 F. L. Chen, T. Y. Ou. Sales forecasting system based on gray extreme learning machine with Taguchi method in retail industry[J]. *Expert Systems with Applications*, 2011, **38**(3): 1336~1445
- 13 Y. Lan, Y. C. Soh, G. B. Huang. Two-stage extreme learning machine for regression[J]. *Neurocomputing*, 2010, **73**(16-18): 3028~3038
- 14 Y. B. Yuan, Y. G. Wang, F. L. Cao. Optimization approximation solution for regression problem based on extreme learning machine[J]. *Neurocomputing*, 2011, **74**(16): 2475~2482
- 15 Z. L. Sun, T. M. Choi, K. F. Au *et al.*. Sales forecasting using extreme learning machine with applications in fashion retailing[J]. *Decision Support System*, 2008, **46**(1): 411~419
- 16 G. B. Huang, X. J. Ding, H. M. Zhou. Optimization method based extreme learning machine for classification[J]. *Neurocomputing*, 2010, **74**(1-3): 155~163
- 17 A. A. Mohammed, R. Minhas, Q. M. Jonathan Wu *et al.*. Human face recognition based on multidimension PCA and extreme learning machine[J]. *Pattern Recognition*, 2011, **44**(10,11): 2588~2597
- 18 W. W. Zong, G. B. Huang. Face recognition based on extreme learning machine[J]. *Neurocomputing*, 2011, **74**(16): 2541~2551
- 19 R. Leardi. Application of genetic algorithm-PLS for feature selection in spectral data sets[J]. *J. Chemometrics*, 2000, **14**(6): 643~655
- 20 R. Leardi, M. B. Seasholtz, R. J. Pell. Variable selection for multivariate calibration using a genetic algorithm: prediction of additive concentrations in polymer films from Fourier transform-infrared spectral data[J]. *Anal. Chim. Acta*, 2002, **461**(2): 189~200
- 21 Shi Jiyong, Zou Xiaobo, Zhao Jiewen *et al.*. Selection of wavelength for strawberry NIR spectroscopy based on BiPLS combined with SAA[J]. *J. Infrared & Millimeter. Waves*, 2011, **30**(5): 458~462
石吉勇, 邹小波, 赵杰文 等. BiPLS 结合模拟退火算法的近红外光谱特征波长选择研究[J]. 红外与毫米波学报, 2011, **30**(5): 458~462
- 22 Xue Long, Li Jing, Liu Muhua *et al.*. Nondestructive detection of soluble solids content on navel orange with Vis/NIR based on genetic algorithm[J]. *Laser & Optoelectronics Progress*, 2010, **47**(12): 123001
薛 龙, 黎 静, 刘木华 等. 基于遗传算法的脐橙可溶性固形物的可见/近红外光谱无损检测[J]. 激光与光电子学进展, 2010, **47**(12): 123001
- 23 Guo Zhiming, Chen Liping, Huang Wenqian *et al.*. Application of genetic algorithm-least squares support vector regression with near infrared spectroscopy for prediction of nicotine content in tobacco[J]. *Laser & Optoelectronics Progress*, 2012, **49**(2): 021201
郭志明, 陈立平, 黄文倩 等. 近红外光谱结合 GA-LSSVR 分析烟草尼古丁含量[J]. 激光与光电子学进展, 2012, **49**(2): 021201