

文章编号: 1001 - 9014(2009)05 - 0386 - 06

基于 iPLS 原理最优化信息区间的桃糖度组合权重 PLS 模型研究

王加华, 李鹏飞, 曹楠宁, 韩东海

(中国农业大学 食品科学与营养工程学院, 北京 100083)

摘要:采用反向区间偏最小二乘法和组合区间偏最小二乘法优化桃糖度可见/近红外光谱的信息区间组合,在选择的信息区间的基础上建立了一种线性组合权重 PLS 模型.对近红外光谱进行二阶导数处理、卷积平滑校正后,发现在区间分割数为 15 时筛选结果最优, BiPLS 所选择的信息区间为 742 ~ 770 nm 和 862 ~ 920 nm, SiPLS 所选择的信息区间为 742 ~ 770 nm、832 ~ 860 nm 和 892 ~ 920 nm.直接组合信息区间 BiPLS 和 SiPLS 模型的 RMSEP 值分别为 0.386 和 0.308,线性组合权重 PLS 模型的 RMSEP 值分别为 0.351 和 0.364.结果说明在近红外定量分析中线性组合权重模型的建立克服了复杂样品各信息区间对 PLS 建模贡献率不一样的问题.

关键词:可见/近红外光谱;区间偏最小二乘法;组合权重 PLS 模型;糖度
中图分类号: O657.3; S123 **文献标识码:** A

STUDY ON THE COMBINATION WEIGHT PLS MODEL FOR DETERMINING SSC OF PEACH BASED ON THE OPTIMAL INFORMATION REGIONS OBTAINED FROM iPLS METHODS

WANG Jia-Hua, LI Peng-Fei, CAO Nan-Ning, HAN Dong-Hai

(College of Food Science & Nutritional Engineering, China Agricultural University, Beijing 100083, China)

Abstract: Backward interval partial least squares (BiPLS) and synergy interval partial least squares (SiPLS) were proposed to search for an optimized combination of information spectral intervals about soluble solids content (SSC) from Vis/NIR spectra of peach. A linear combination weight PLS model was developed on the basis of the selected information intervals. The spectra were preprocessed by second-order derivative and Savitzky-Golay smoothing. It is found that the selected result is the best when the interval number is 15. The information intervals selected by BiPLS are 742 ~ 770 nm and 862 ~ 920 nm, while those selected by SiPLS are 742 ~ 770 nm, 832 ~ 860 nm and 892 ~ 920 nm. For BiPLS and SiPLS models of direct combination intervals, the root mean square error of prediction (RMSEP) are 0.386 and 0.308, respectively. And for the PLS models of linear combination weight, the RMSEP are 0.351 and 0.364, respectively. The results reveal that the proposed method overcomes the difficulties that the different information intervals for the complicated samples have different contribution to PLS models. The method is very promising for vibrational spectroscopy and it gives much better prediction than the whole-spectrum PLS modeling.

Key words: visible/near infrared spectrum; interval partial least squares (iPLS); combination weight PLS model; soluble solids content (SSC)

引言

桃 (*Prunus persica*) 原产于中国,近年来其产量和栽培面积都居世界首位.桃的品质主要分为感官品质和营养品质,是由产品外观和众多内在因素的复合评价因子构成的,糖度是评价桃内部品质的重

要指标之一^[1].因此,需要一种快速无损技术进行糖度检测,在桃商品化中实现按品质分级,以保证品质的一致性.

近红外光谱结合化学计量学方法已被广泛应用于水果品质分析^[2].偏最小二乘法 (PLS) 具有良好的处理光谱数据共线性能力,常用于解析近红外光

谱信息^[3,4].在建立 PLS模型时,常进行信息区间选择以简化模型和提高模型精度,如遗传算法^[5]、交互信息变量选择^[6]、移动窗口偏小二乘法^[7]、区间偏小二乘法(iPLS)^[8]等及其它们的演变.其中基于 iPLS原理的扩展方法,如正向区间偏小二乘法(FiPLS)^[9]、反向区间偏小二乘法(BiPLS)^[9]和组合区间偏小二乘法(SiPLS)^[10]已被广泛研究和应用.对于多个被选择信息区间,以往研究主要采用合并所有信息区间建立 PLS模型^[9,10],然而信息区间所承载的信息变量不同,也就是对模型贡献率不一样,需要在各信息区间建立独立的 PLS模型,然后建立一个线性组合权重模型^[7],从而获得更高预测精度.

采用可见/近红外漫透射技术,通过反向区间偏小二乘法和组合区间偏小二乘法优化桃糖度的信息区间组合,建立一种线性组合权重 PLS模型.以交互验证均方根误差考察信息区间最优化效果,以预测均方根误差和预测回收率来评价模型预测精度.

1 原理和算法

1.1 BiPLS和 SiPLS

iPLS的原理是,首先将整个光谱分割成 k 个等宽子区间,然后在每个子区间进行偏小二乘回归.采用留一交互验证法(LOO-CV)计算交互验证均方根误差(RMSECV),当 RMSECV 值最小时,对应的因子数为最佳因子数,采用最佳因子数在各子区间建立局部最优 PLS模型.

BiPLS是在 iPLS的基础上,依次减少信息量最差或共线性变量最多的 $i(i=0, 1, 2, \dots, k)$ 个区间,即去除 RMSECV 值最大的区间,在剩余的 $k-i$ 个区间上建立最优 PLS模型,并给出相应的 RMSECV 值.当 RMSECV 最小时所对应的多个区间即为所优化的组合区间. SiPLS是在各子区间上,计算所有可能的 $j(2 \leq j \leq k)$ 个子区间组合模型,依 RMSECV 值大小给出各个组合区间,选择 RMSECV 最小时对应的组合区间.

1.2 线性组合权重模型

当多个信息区间被选择时,在每个信息区间建立局部最优 PLS模型,对各 PLS模型进行线性组合,得到一个组合权重模型.组合权重采用最小二乘法计算,即样品的测量值和组合模型的预测值之间偏差的平方和 $\|c - y\|^2$ 最小,组合权重根据式(1)计算:

$$\begin{aligned} \text{minimize} & \left\| c - \sum_{j=1}^J w_j R_j b_j \right\|^2, \\ \text{subject to} & 1 - w_j \geq 0, \sum_{j=1}^J w_j = 1 (j = 1, 2, \dots, J) \end{aligned} \quad (1)$$

其中, c 是样品真实浓度矩阵; w_j 是第 j 个 PLS模型权重; R_j 是第 j 个信息区间光谱矩阵; b_j 是第 j 个 PLS模型回归系数; J 是所选择的信息区间数.

组合权重 (w) 反映各信息区间的 PLS模型对预测结果的贡献率, w 大则贡献率大,表明该信息区间承载信息变量多.在直接使用式(1)计算各局部模型权重时必须满足一个条件:局部模型的建立所采用的校正集和预测集应有代表性,各模型均具有最优的预测精度.本文最终预测结果采用各局部模型预测值的加权计算获得,预测结果 (Y_p) 可表述为:

$$Y_p = \sum_{j=1}^J w_j Y_j, \quad (2)$$

其中, Y_j 是第 j 个 PLS模型预测值.

线性组合模型的模型精度评价指标为 RMSEP, 计算公式如下:

$$\text{RMSEP} = \sqrt{(c - Y_p)^T (c - Y_p) / m}, \quad (3)$$

其中, m 是预测集样品数.

1.3 预测回收率

预测回收率是指样品预测值与真实值的比率,表示预测值偏离理想预测回归线的程度.模型预测精度越高,适用性越好,其预测回收率越接近 1. 预测回收率计算公式如下:

$$T = \frac{y_{pi}}{y_{0i}} \times 100\%, \quad (4)$$

其中, y_{0i} 是预测集 i 样品的真实值; y_{pi} 是预测集 i 样品的模型计算值.

2 实验部分

2.1 仪器与样品

便携式可见/近红外光谱仪 K-BA100R型 (Kubota, Japan) 配备光纤采集附件,采用岛津线阵检测器(256像元).手持数字式糖度仪 PAL-1型 (Atago, Japan) 测量桃糖度,读数结果为 Brix,具备自动温度校正功能. 86个桃样品购自北京某批发市场.

2.2 光谱采集与标准值测定

样品恒温至室温(28℃)后,手动放置于光纤头上,采用漫透射方式采集光谱,以内置分光作为背景.样品和背景积分时间均为 50ms,光谱采集范围为 500~1010nm,间隔为 2nm,共有 256个数据点.

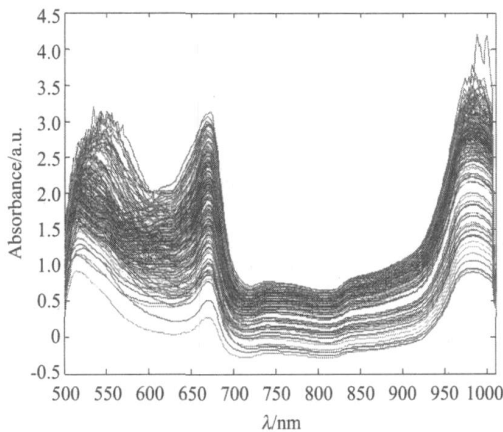


图1 桃可见/近红外漫透射原始光谱图
Fig.1 Original Vis/NIR spectra of peach

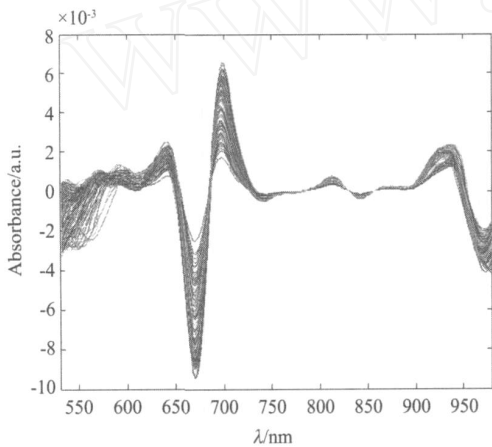


图2 经二阶导和卷积平滑处理后的光谱
Fig.2 Spectra after preprocessed by second-order derivative and Savitzky-Golay filter smoothing

采集5次,将5次的平均光谱作为样品光谱.每个桃以缝合线为界,分别采集两面光谱,共获得172条光谱,如图1所示.采集光谱后,在采集面上取直径约40~50mm的果肉测定标准值.

2.3 数据处理软件

区间偏最小二乘法工具包由Nørgaard等提供的网络共享获得,各计算程序自行编写,在MATLAB 7.0 (The mathworks Inc., Natick, MA)中实现.光谱预处理方法在TQ 6.2 (Themo Nicolet, USA)中实现.

3 结果与讨论

3.1 光谱数据预处理

原始光谱图在整个光谱区500~1010nm存在光谱基线漂移,原始光谱中有两个明显吸收峰,在680nm附近为叶绿素吸收峰,在960nm附近为水的

表1 校正集和预测集样品标准值统计表

Table 1 Statistics of SSC in peach of calibration and prediction data sets (°Brix)

Sample set	Number of samples	Range	Average	Standard deviation
calibration	130	6.3~12.0	8.93	1.17
prediction	42	6.8~11.1	8.90	1.09

表2 不同区间数下的BiPLS优化结果

Table 2 Optimization result of BiPLS for different numbers of interval

Number of interval	Selected intervals	Spectral range/nm	RMSECV
10	[8 9 4 6]	670~714, 762~804, 850~936	0.410
15	[8 13 12]	742~770, 862~920	0.370
20	[17 11 6 16 12]	652~672, 762~804, 872~914	0.393
25	[20 13 21 8]	658~674, 748~764, 874~908	0.407

吸收峰.为了消除基线漂移和放大特征信息,采用二阶导数处理,但是在放大信息的同时,噪声也被放大,为了消除噪声影响,对二阶导数光谱进行Savitzky-Golay二项式卷积平滑,平滑窗口大小为25.处理后光谱去掉头16和尾15个信噪比差的数据点,剩余225个数据点用于下一步处理.处理后的光谱如图2所示,在700~900nm范围内信息更加丰富凸显.

所有样本按照浓度梯度排序,依3:1左右分为校正集和预测集,测量值的最大和最小样本归为校正集,统计参数如表1所示.

3.2 BiPLS选择信息区间

近红外光谱既表征目标信息,同时也受到非目标信息或仪器噪声干扰,而且目标信息区间分布复杂,信息区间宽度不等.因此,在运行BiPLS选择信息区间时,应考察区间分割数对选择结果及模型的影响.在PLS计算中,最大因子数应该不大于所包含的变量数,在运行程序时所有参数设置相同,预处理方法选用自动多元散射校正(MSCAUTO),采用LOO-CV法计算RMSECV,将整条光谱(225个数据点)分为 k 个子区间($k=10, 15, 20, 25$),最大因子数设为8,在设定子区间数下,最优的BiPLS运行结果如表2所示.

由表2可得,当模型具有最小的RMSECV值时,对应的子区间间隔数为15,此时所选择的波长范围是742~770nm、862~920nm,数据点数为45个.

3.3 SiPLS选择信息区间

由BiPLS运算结果可知,当光谱被分割成15个

表 3 不同组合数下的 SiPLS优化结果

Table 3 Optimization results of SiPLS for different numbers of combination

Number of combination	Selected intervals	Spectral range/nm	RMSECV
5	[3 5 8 12 13]	592 ~ 620, 652 ~ 680, 742 ~ 770, 862 ~ 920	0.397
4	[8 10 12 13]	742 ~ 770, 802 ~ 830, 862 ~ 920	0.373
3	[8 11 13]	742 ~ 770, 832 ~ 860, 892 ~ 920	0.365
2	[8 13]	742 ~ 770, 892 ~ 920	0.378

子区间时,可获得最佳组合区间。因此,在使用 SiPLS程序时,也将光谱分为 15个子区间,每个区间 15个数据点。在使用 SiPLS选择最佳组合区间时,不同的组合数对模型有一定影响,即引入变量时可能会提高模型精度,也可能会引入噪声降低模型精度。在本文中,选择最佳组合区间时,对比了不同组合数下的最优结果,如表 3所示。

从表 3可以看出,当模型具有最小的 RMSECV 值时,对应的组合数为 3,所选择的光谱区间为 742 ~ 770nm、832 ~ 860nm、892 ~ 920nm,数据点数为 45 个。

图 3为采用 iPLS法计算在子区间所见独立 PLS模型结果,虚线为全谱最佳因子数下的 RMSECV,斜体数字为各子区间 PLS模型最佳因子数。从图中可以看出较小的 RMSECV 所对应的区间依次为:742 ~ 770nm、862 ~ 890nm、892 ~ 920nm和 832 ~ 860nm。BiPLS所选择的最优组合区间为 742 ~ 770nm和 862 ~ 920nm,与 iPLS结果一致,分别对应

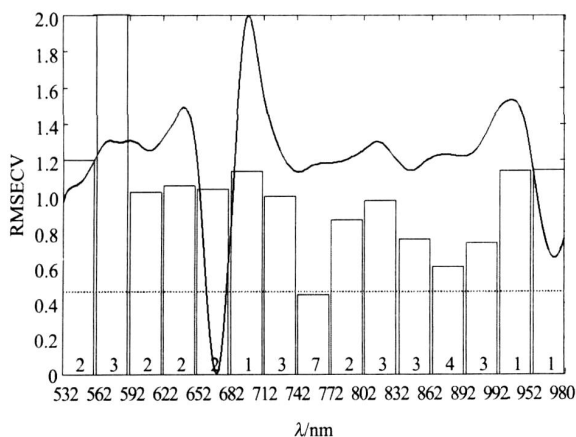


图 3 iPLS优化结果图(区间分割数为 15)

Fig 3 Location of the optimum intervals corresponding to full spectrum obtained by iPLS(Dotted line is RMSECV (8LV 's) for global model/ Italic numbers are optimal LVs in interval model)

iPLS中 RMSECV 最小的 3个子区间。SiPLS所选择的组合为:742 ~ 770nm、832 ~ 860nm和 892 ~ 920nm,与 BiPLS结果有一定差别,主要原因是 SiPLS具有较好的组合性,是在对全谱信息分解的基础上再进行重构,所得到的组合区间并不是 iPLS法最优的几个区间简单的线性相加。

3.4 组合权重模型建立

在建立最终线性组合模型,直接使用式(1)计算组合权重时,要求在单个信息区间所建立的独立模型具有最佳预测精度,一般的方法是以 RMSEP来衡量。在被选择区间内建立最优 PLS模型,其结果如表 4所示。

对于 BiPLS所选择的最优组合区间为 742 ~ 770nm和 862 ~ 920nm,采用式(1)分别计算组合权重为 0.417和 0.583,采用式(3)计算 RMSEP值,其组合权重模型(BiPLScom)结果如表 4所示。对 SiPLS所选择区间 742 ~ 770nm、832 ~ 860nm和 892 ~ 920nm,同样采用式(1)分别计算组合权重为 0.459、0.173和 0.368,其组合权重模型(SiPLScom)结果如表 4所示。所计算权重值与表 4中独立 iPLS模型的 RMSEC值存在相关性,RMSEC小权重较大,对应区间承载的有效信息多,该 PLS模型对组合权重 PLS模型贡献大。

表 4显示所有 iPLS模型的 RMSEP小于等于 0.517,均优于全谱 PLS模型的 0.883。经 BiPLS和 SiPLS优化信息区间后的模型精度进一步提高,其直接组合信息区间的 PLS模型 RMSEP分别为 0.386和 0.308。经 BiPLS和 SiPLS优化信息区间后,建立组合权重 PLS模型,BiPLScom模型 RMSEP值相对直接组合模型有所降低;但是 SiPLScom模

表 4 在给定信息区间下不同方法所建桃糖度模型结果统计表

Table 4 Results of PLS models for SSC of peach using given spectral regions and evaluation of the performance of the their corresponding PLS models

Model	Spectral range/nm	Latent variables	R	RMSEC	RMSEP
PLS	532 ~ 980	8	0.767	0.751	0.883
iPLS1	742 ~ 770	5	0.939	0.396	0.440
iPLS2	832 ~ 860	6	0.859	0.592	0.517
iPLS3	892 ~ 920	5	0.932	0.418	0.504
iPLS4	862 ~ 920	5	0.946	0.376	0.402
BiPLS	742 ~ 770, 862 ~ 920	7	0.945	0.377	0.386
BiPLScom	742 ~ 770, 862 ~ 920	-	0.952	0.354	0.351
SiPLS	742 ~ 770, 832 ~ 860, 892 ~ 920	8	0.957	0.336	0.308
SiPLScom	742 ~ 770, 832 ~ 860, 892 ~ 920	-	0.954	0.354	0.364

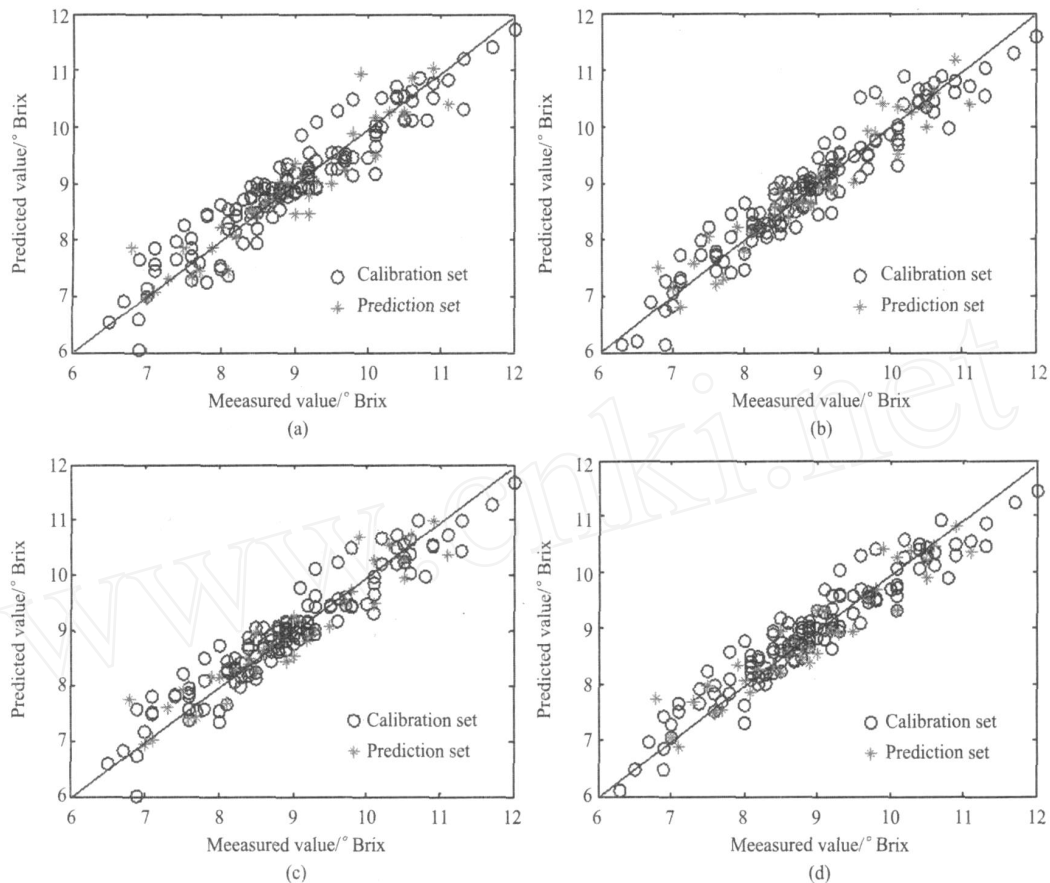


图 4 不同 PLS模型预测结果图 (a) BiPLS (b) SiPLS (c) BiPLScm (d) SiPLScm

Fig 4 Plots of measured SSC values vs N IR predicted values (a) BiPLS model (b) SiPLS model (c) BiPLScm model (d) SiPLScm model

型 RMSEP值相对直接组合模型却变大,主要是 SiPLS 算法在优化组合中排除了共线性变量,保留的变量能够极好地表征目标信息,在组合权重模型中所用变量较多,存在共线性问题,导致预测结果变差。

3.5 预测结果

经二阶导数、Savitzky-Gp lay卷积平滑处理后,在采用 BiPLS和 SiPLS法所选择的组合区间,分别建立 PLS模型并进行预测。直接组合信息区间的 BiPLS和 SiPLS模型预测结果如图 4 (a)、(b),预测集的 r 值分别为 0.937和 0.960, RMSEP值分别为 0.386和 0.308,根据式 (4)计算预测回收率范围分别为 93.74% ~ 115.52%和 93.77% ~ 110.09%;线性组合权重的 BiPLScm和 SiPLScm模型的预测结果如图 3 (c)、(d),预测集的 r 值分别为 0.948和 0.947, RMSEP值分别为 0.351和 0.364,根据式 (4)计算预测回收率范围分别为 93.39 ~ 114.07%和 92.33 ~ 113.66%。

4 结论

采用可见/近红外漫透射技术,通过反向区间偏最小二乘法和组合区间偏最小二乘法优化桃糖度的信息区间组合,建立了一种线性组合权重 PLS模型,克服了复杂样品各信息区间对 PLS建模贡献率不同的问题。结果表明:当光谱经二阶导和卷积平滑后,区间分割数为 15时, BiPLS所选择的信息区间为 742 ~ 770nm和 862 ~ 920nm, SiPLS所选择的信息区间为 742 ~ 770nm、832 ~ 860nm和 892 ~ 920nm;直接组合信息区间 BiPLS和 SiPLS模型的 RMSEP值分别为 0.386和 0.308;线性组合权重 PLS模型的 RMSEP值分别为 0.351和 0.364。以 RMSEP和预测回收率来评价模型预测精度,直接组合信息区间建立 SiPLS模型获得最佳预测精度,其 RMSEP值为 0.308,预测回收率范围为 93.77% ~ 110.09%。

REFERENCES

- [1] ZHANG Hai-Ying, HAN Tao, WANG You-Nian, *et al* Selection of factors for evaluating peach (*Prunus persica*) fruit quality[J]. *Transactions of the CSAE* (张海英,韩涛,

- 王有年,等. 桃果实品质评价因子的选择. 农业工程学报), 2006, 22(8): 235—239.
- [2] HAN Dong-Hai, WANG Jia-Hua. Review of nondestructive measurement of fruit quality by means of near infrared spectroscopy[J]. *Chinese Journal of Lasers* (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. 中国激光), 2008, 35(8): 1123—1131.
- [3] LIU Yan-De, LUO Ji, CHEN Xing-Miao. Analysis of soluble solid content in nanfeng mandarin fruit with visible near infrared spectroscopy[J]. *J. Infrared Millim. Waves* (刘燕德, 罗吉, 陈新苗. 可见近红外光谱的南丰蜜桔可溶性固形物含量定量分析. 红外与毫米波学报), 2008, 27(2): 119—122.
- [4] NI Li-Xing, WANG Gang-Li, LI Zhi-Meng, et al. Qualitative and quantitative analysis of Tongren Wuji Baifeng pills by near infrared spectroscopy[J]. *J. Infrared Millim. Waves* (聂黎行, 王钢力, 李志猛, 等. 近红外光谱法对同仁乌鸡白凤丸的定性和定量分析. 红外与毫米波学报), 2008, 27(3): 205—210.
- [5] Ying Yi-bin, Liu Yan-de. Nondestructive measurement of internal quality in pear using genetic algorithms and FT-NIR spectroscopy[J]. *Journal of Food Engineering*, 2008, 84: 206—213.
- [6] Durand A, Devos O, Ruckebusch C, et al. Genetic algorithm optimization combined with partial least squares regression and mutual information variable selection procedures in near-infrared quantitative analysis of cotton viscose textiles[J]. *Analytica Chimica Acta*, 2007, 595: 72—79.
- [7] Jiang J H, Berry R J, Ozaki Y, et al. Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data [J]. *Anal. Chem.*, 2002, 74: 3555—3565.
- [8] Nørgaard L, Saudland A, Wagner J, et al. Interval partial least squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy [J]. *Applied Spectroscopy*, 2000, 54: 413—419.
- [9] Zou Xiao-bo, Zhao Jie-wen, Li Yan-xiao. Selection of the efficient wavelength regions in FT-NIR spectroscopy for determination of SSC of Fuji' apple based on BiPLS and FiPLS models [J]. *Vibrational Spectroscopy*, 2007, 44: 220—227.
- [10] ZHU Xiang-Rong, LI Na, SHI Xin-Yuan, et al. Analysis of geniposide and total nitrogen content in Qingkailing injection intermediate by near infrared spectroscopy and synergy interval partial least squares[J]. *Chin. J. Chinese Universities* (朱向荣, 李娜, 史新元, 等. 近红外光谱与组合的间隔偏最小二乘法测定清开灵四混液中总氮和栀子苷的含量. 高等学校化学学报), 2008, 29(5): 906—911.
- (上接 356页)
- tra[J]. *China Science Bulletin* (刘沐华, 张学工, 孙素琴. 中药材产地的近红外光谱自动鉴别和特征谱段选择. 科学通报), 2005, 50(4): 393—398.
- [5] WU Yong-Jun, LI Wei, Xiang Bing-Ren, et al. Identification of traditional Chinese medicine Baizhi with near-infrared spectrum[J]. *Journal of Chinese Medicinal Materials* (吴拥军, 李伟, 相秉仁, 等. 近红外光谱技术用于白芷类中药的鉴定研究. 中药材), 2001, 24(1): 26—28.
- [6] LIU Gou-Lin, CAI Jin-Na, LI Wei, et al. Near-Infrared spectroscopy technique used in the classification of the *Cnidium Monnieri* (L.) Cusson [J]. *Computer and Applied Chemistry* (刘国林, 蔡金娜, 李伟, 等. 近红外光谱技术在中药蛇床子分类中的应用. 计算机与应用化学), 2000, 17(2): 109—110.
- [7] Delwiche S R, Bean M M, Miller R E, et al. Apparent amylose content of milled rice by near-infrared reflectance spectrophotometry[J]. *Cereal Chemistry*, 1995, 72: 182—187.
- [8] Bao J S, Cai Y Z, Corke H. Prediction of rice starch quality parameters by near-infrared reflectance spectroscopy [J]. *Food Science*, 2001, 66: 936—939.
- [9] Wu J G, Shi C H, Zhang X M. Estimating the amino acid composition in the milled rice powder by near-infrared reflectance spectroscopy[J]. *Field Crops Research*, 2002, 75: 1—7.
- [10] WANG Hai-Lian, WAN Xiang-Yuan, HU Pei-Song, et al. Quantitative analysis of fat content in brown rice by near infrared spectroscopy (NIRS) technique[J]. *Scientia Agricultura Sinica* (王海莲, 万向元, 胡培松, 等. 稻米脂肪含量近红外光谱分析技术研究. 中国农业科学), 2005, 38(8): 1540—1546.
- [11] BI Jing-Cui, ZHANG Wen-Wei, XIAO Ying-Hui, et al. Analysis for protein content in rice by near infrared reflectance spectroscopy (NIRS) technique[J]. *Acta Agronomica Sinica* (毕京翠, 张文伟, 肖应辉, 等. 应用近红外光谱技术分析稻米蛋白质含量. 作物学报), 2006, 32(15): 709—715.
- [12] SHAO Yong-Ni, CAO Fang, HE Yong. Discrimination years of rough rice by using visible/near infrared spectroscopy based on independent component analysis and BP neural network [J]. *J. Infrared Millim. Waves* (邵咏妮, 曹芳, 何勇. 基于独立组分析法和BP神经网络的可见/近红外光谱稻谷年份的鉴别. 红外与毫米波学报), 2007, 26(6): 433—436.
- [13] Haykin S. *Neural network—a comprehensive foundation* [M]. New York, USA: Macmillan College Publishing Company, 1994, 1—44.
- [14] QI Xiao-Ming, ZHANG Lu-Da, DU Xiao-Lin, et al. Quantitative analysis using NIR by building PLS-BP model [J]. *Spectroscopy and Spectral Analysis* (齐小明, 张录达, 杜晓林, 等. PLS-BP法近红外光谱定量分析研究. 光谱学与光谱分析), 2003, 23(5): 870—872.
- [15] Workman J. *Handbook of Organic Compounds: NIR, IR, Raman, and UV-VIS Spectra Featuring Polymers and Surfactants* [M]. San Diego, USA: Academic Press, 2001: 77—197.
- [16] Sasic S, Ozaki Y. Short-wave near-infrared spectroscopy of biological fluids. I In Quantitative analysis of fat, protein, and lactose in raw milk by partial least-squares regression and band assignment [J]. *Anal. Chem.*, 2001, 73: 64—71.