# 基于粒子群算法的波长选择方法用于苹果酸度的近红外光谱分析

夏阿林*[1]，叶华俊[2]，周新奇[2]，王 健[1]
(1. 杭州电子科技大学电子信息学院 杭州 310018; 2. 聚光科技(杭州)有限公司，杭州 310052)

摘 要: 采用便携式近红外光谱分析仪，对苹果样品进行扫描获得光谱数据，运用偏最小二乘法结合基于粒子群算法的波长选择方法对苹果试验数据进行多元统计分析，建立数学模型，利用该模型对苹果酸度进行了预测。对于基于粒子群算法和全谱偏最小二乘方法，校正集样品的酸度预测值和实测值之间的相关系数分别为 0.9880和 0.9553，校正均方根误差分别为 0.0197和 0.0388；预测集样品的酸度预测值和实测值之间的相关系数分别为 0.9833和 0.9596，预测均方根误差分别为 0.0193和 0.0304。与全谱偏最小二乘法相比，基于粒子群算法的偏最小二乘法，不仅较大地减少波长变量而降低计算量，而且也较大地提高了模型性能而增强了模型预测的准确性。该方法可建立较好的定量分析模型，能广泛应用于现场或野外苹果酸度的快速分析。

关键词: 近红外；粒子群；苹果；酸度
中图分类号: O657.3 文献标识码: A 文章编号: 1000-0720(2010)09-012-04

随着人们生活水平的提高，消费者对于水果品质的要求也在不断提高，如酸度、糖度和营养成分等内部品质信息越来越受到人们的关注[1]。近红外光谱分析技术无需对样品作任何化学和物理的预处理，即可获取样品内部深处的物质信息，与传统的化学分析及其它光谱分析方法相比，具有价廉、方便、快速和无损等特点[2]。

然而近红外区的谱带复杂、重叠多，通过特定方法选取特征变量有可能得到更好的定量校正模型。波长选择一方面可以简化模型，更主要的是由于不相关或非线性变量的剔除，可以得到预测能力更强的校正模型。在多元校正分析中，波长选择方法主要有相关系数法、方差分析法、逐步回归法、无信息变量消除法、间隔偏最小二乘法、遗传算法、模拟退火、粒子群算法(PSO)等[3-7]。粒子群算法自 90年代提出以来，已在许多领域得到应用[7-9]，但在近红外光谱波长变量选取方面，国内外报道较少[10,11]。

本文将粒子群算法用于苹果酸度的近红外光谱分析中的波长选择，选择后的波长变量再由偏最小二乘(PLS)方法建立分析校正模型，与全光谱偏最小二乘(W-PLS)相比，该方法较大地提高了模型的预测能力。

## 1 基本原理

### 1.1 粒子群算法

粒子群算法最早是 1995 年由 Kennedy和 Eberhart受人工生命的研究结果启发而提出的一种全局随机优化技术[7]，其基本概念源于对鸟群捕食行为的研究，是一种源于对鸟群捕食行为研究的进化计算技术，通过粒子间相互作用发现复杂搜索空间中的最优区域。类似于遗传算法，它是一种基于种群的全局优化技术，系统初始化一组随机解，通过迭代找到最优值，但是并没有遗传算法用的交叉以及变异，在解空间追随最优的

* 收稿日期: 2009-12-25; 修订日期: 2010-03-20
基金项目: 国家 863项目(2009AA04Z129)和浙江省重大应用电子技术和新型电子元器件专项(2007C11091)资助
作者简介: 夏阿林 (1974-)，男，讲师，E-mail: alinxia@hdu.edu.cn

粒子进行搜索。与遗传算法相比，粒子群算法的优势在于概念简明，容易实现，同时又有深刻的智能背景，并且没有过多的参数需要调整。随机初始化群体中每个粒子的位置和速度，使它们分散在整个空间中。每个粒子是 D维空间的一个点。第 $i$个粒子表示为一个 D维向量 $x_i = (x_{i1}, x_{i2}, \cdots, x_{iD})$；第 $i$个粒子的“飞翔”速度，即第 $i$个粒子位置变化的速率表示为 $v_i = (v_{i1}, v_{i2}, \cdots, v_{iD})$。算法假定所有的粒子都朝着个体最优位置和全局最优位置移动，记第 $i$个粒子迄今为止搜索到的最优位置即个体最优位置为 $p_i = (p_{i1}, p_{i2}, \cdots, p_{iD})$，整个粒子群迄今为止搜索到的最优位置为全局最优位置 $p = (p_{g1}, p_{g2}, \cdots, p_{gD})$。在每次迭代中，粒子通过跟踪这两个最优值来更新自己，即粒子本身所找到的最优值和整个种群目前为止找到的最优值。在找到这两个最优值时，粒子根据如下的公式来更新自己的速度和位置:

$$v_{id}(new) = w * v_{id}(old) + c_1 * r_1 * (p_{id} - x_{id}) + c_2 * r_2 * (p_{gd} - x_{id}) \quad (1)$$

$$x_{id}(new) = x_{id}(old) + \mu * v_{id}(new) \quad (2)$$

这里，$w$是非负常数，称为惯性因子，用以平衡全局搜索与局部搜索，这里取 $w = 0.5$；学习因子 $c_1$和 $c_2$是非负常数，根据经验这两个常数都取整数值 2；$r_1$和 $r_2$为介于 [0, 1]之间的随机数；$\mu$称为约束因子，是控制速度的权重。首先由 (1)式根据粒子原来的速度和粒子目前位置与两个最优位置的距离来计算粒子的新速度。然后粒子根据式 (2)移动到一个新的位置。粒子就是这样通过跟踪两个“极值”来更新自己。如果达到了规定最小误差标准或迭代达到了规定次数，程序终止。为了避免 PSO算法收敛于局部最优，增强算法克服局部最优的能力，迫使 10% 的粒子随机飞行，不追随两个最优值。

1.2 波长选择

基于粒子群算法的波长选择采用二进制编码方式。每个粒子长度等于全部的波长点数，每个波长对应一个二进制码，其中数值 1表示所对应的波长被选中，而数值 0表示所对应的波长未被选中。粒子在每一维上的移动被限定为 0或是 1，即在二进制问题中，更新一个粒子则表示变化 D维空间一个元素为 0或是 1，速度表示元素 $x_{id}$取值 0或是 1的几率。粒子的适应度函数为偏最小二乘交互验证中苹果酸度预测值和实际值的均方差。交互验证均方根误差 (RMSECV)的计算如下:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{K_p} (\tilde{c}_i - c_i)^2}{K_p}}$$

其中，$K_p$表示交互验证集样本数；$\tilde{c}_i$表示第 $i$个样本的预测浓度；$c_i$表示第 $i$个样本的实测浓度。算法首先随机初始化 40个体，算法循环 80代。

## 2 实验部分

### 2.1 样品及参考值分析

选取无明显外部缺陷，颜色较均匀的红富士苹果 68个，秦冠苹果 61个，随机将它们分别编号后置于 4 ℃冰柜中贮藏。光谱检测实验在室温下 (26 ℃) 进行。实验前，将冰柜中取出的苹果置于试验室中 4h，以使苹果整体温度达到与环境温度的一致。苹果样品的光谱采集与酸度化学值的测定当天完成，酸度的测定参照国标 GB/12293—1990，采用指示剂滴定法，苹果酸度的化学值统计结果如表 1所示，可见样品具有较好代表性，分布范围较宽。

表 1 苹果酸度化学值的统计结果

**Tab. 1 Statistic of apple acidity**

| | 样品数 | 最小值 | 最大值 | 极差 | 平均值 | 标准偏差 |
|---|---|---|---|---|---|---|
| 校正集 | 68 | 0.1869 | 0.8451 | 0.6582 | 0.3804 | 0.1270 |
| 预测集 | 61 | 0.2119 | 0.6730 | 0.4611 | 0.4370 | 0.1060 |

## 2.2 仪器及光谱采集

SupNIR1100便携式近红外光谱分析仪 (聚光科技)[12], 采用硅阵列检测器, 配有近红外光纤附件, 光谱范围 600~ 1100 nm, 分辨率 6 nm, 波长间隔 1 nm, 平均次数为 3次。以陶瓷白板作为参比材料, 测量方式为漫反射。在每个苹果的最大横径上进行光谱扫描, 129个苹果的近红外吸光度光谱如图 1所示。

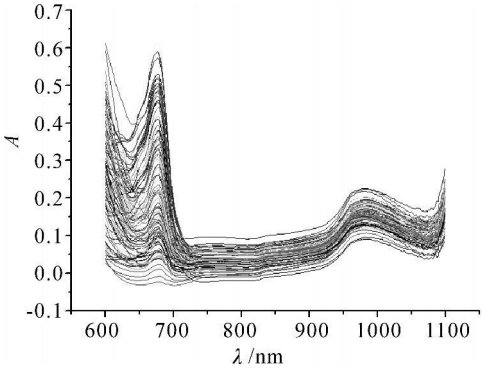


图 1 苹果的近红外吸光度光谱

**ig 1 Original absorbance spectra of the apple samples**

## 3 结果与讨论

### 3.1 近红外光谱预处理

为了消除高频随机噪声以及样本不均导致的基线漂移的影响, 对苹果样品的原始光谱数据进行预处理。先用 Savitzky-Golay平滑法 (窗口参数 7, 拟合次数 2)消除高频噪声对信号的影响; 再对光谱进行多元散射校正 (MSC), 消除光程及颗粒大小差异引起的基线漂移; 最后对光谱数据作均值中心化处理, 这样处理后的光谱数据充分反映了变化信息, 使所有的数据都分布在零点两侧, 对于以后的回归运算可以简化并使之稳定。

### 3.2 建模及预测

从全部 129个苹果样品中, 随机取 68个作为校正样品, 剩余 61个作为预测样品。建模过程中的最佳主因子数由交互验证法确定。采用 W-PLS与 PSO-PLS方法建立的苹果酸度模型的分析结果如表 2所示。图 2a和图 2b分别是使用 W-PLS和 PSO-PLS方法获得的酸度校正集样品和预测集的化学值 (参考值)与相应预测值的相关曲线, 横轴为化学值, 纵轴为预测值。从图 2可见,

表 2 W-PLS和 PSO-PLS分析苹果酸度的结果

**Tab. 2 The results for W-PLS and PSO-PLS models**

| 建模方法 | 变量数 | 主因子数 | RMSEC | Rc | RMSEP | Rp |
|---|---|---|---|---|---|---|
| W-PLS | 500 | 7 | 0.0388 | 0.9553 | 0.0304 | 0.9596 |
| PSO-PLS | 53 | 6 | 0.0197 | 0.9880 | 0.0193 | 0.9833 |

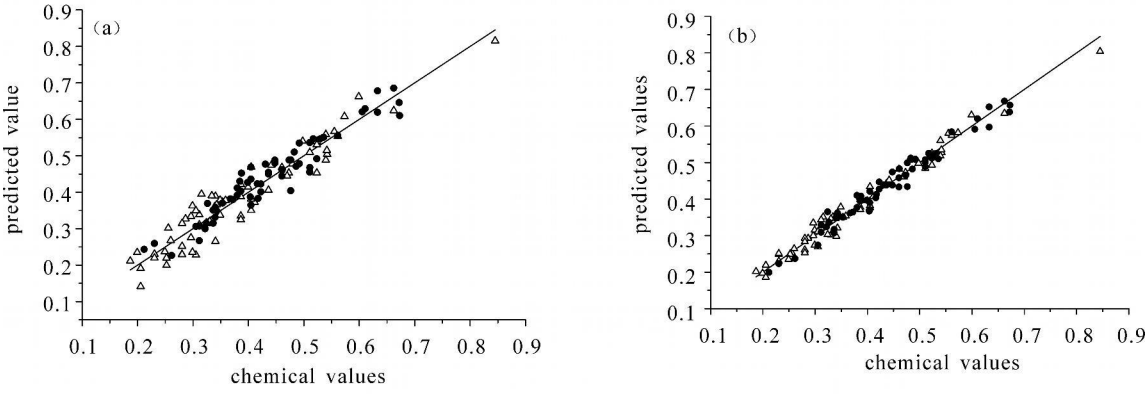


图 2 苹果酸度的化学值与使用全谱 PLS法 (a)及 PSO-PLS法 (b)获得的预测值关系

**Fig 2 Correlation of chemical value and predicted value for acidity of apples by W-PLS (a) and PSO-PLS (b)**

△ calibration set; ● prediction set

PSO PLS方法与 W-PLS相比, 获得的苹果酸度的预测值与化学值之间具有更好的相关性。

从表 2中可见, PSO-PLS方法与 W-PLS相比, 波长变量数由 500个减少到 53个, 可以较大地减少计算量; 校正均方根误差 (RMSEC)明显变小以及校正相关系数 (Rc)变大, 说明模型优化明显; 预测均方根误差 (RMSEP)降低明显以及预测相关系数 (Rp)变大, 表明预测准确度有较大的提高。结果显示, PSO-PLS法所建立的苹果酸度近红外光谱模型比全光谱模型更简洁、更稳健, 该模型具有较强的预测能力。

## 参考文献

[1] 王加华, 韩东海 . 光谱学与光谱分析, 2008, 28(10): 2308
[2] 严衍禄, 赵龙莲, 韩东海等 . 近红外光谱分析基础与应用 . 北京: 中国轻工业出版社, 2005. 1- 6
[3] Rinband C J, Massant D L. Anal Chem, 1995, 67(23): 4295
[4] Lucasius C B, M Bekers M L, Kateman G. Anal Chim Acta, 1994, 286: 135
[5] Kalivas J K, Robents N, Sutter J M. Anal Chem, 1989, 61(18): 2024
[6] Spiegelman C H, Mcshane M H J, M J Goetz *et al*. Anal Chem, 1998, 70(1): 35
[7] Kenned Y J, Eberhart R. Proceedings of IEEE International Conference on Neural Networks. Perth, Australia, 1995, 1942
[8] Shen Q, Jiang J H, Jiao C X *et al*. Eur J Pharm Sci, 2004, 22: 145
[9] 陶丘博, 申 琦, 张小亚等 . 分析化学, 2009, 37(8): 1197
[10] Wang X, Yang C H, Qin B *et al*. J Control Theory Appl, 2005, 4: 371
[11] Liu F, He Y. J Agric Food Chem, 55: 8883
[12] 叶华俊, 刘立鹏, 张学峰等 . 光学技术, 2008, 34(suppl): 66

**Near infrared determination of acidity in applesby wavelength variable selection based on particle swarm optimization algorithm**

*XIA A-lin*[*1], *YE Hua-jun*[2], *ZHOU Xin-qi*[2] and *WANG Jian*[1] (1. Electronic Information College, Hangzhou Dianzi University, Hangzhou 310018, 2. Focused Photonics (Hangzhou), Inc., Hangzhou 310052), Fenxi Shiyanshi, 2010, 29(9): 12~ 15

**Abstract**: The spectrum data were obtained by direct scanning apples with the portable near infrared analyzer. In order to decrease the force of the spectrum noise and sample graininess, raw spectra were pretreated by Savitzky-Golay smoothing and multiplication scatter correction. The establishment of the calibration model was based on partial least square method (PLS) and the model was optimized by wavelength variable selection with particle swarm optimization algorithm (PSO). To illuminate the performance of the optimized method, the PLS method based on swarm optimization algorithm (PSO-PLS) was compared with the PLS method based on the whole spectra (W-PLS) by the analysis of apple acidity. The calibration and prediction acidity models respectively gave the correlation coefficients of 0.9880 and 0.9833 for PSO-PLS and of 0.9553 and 0.9596 for W-PLS; the root mean standard errors of calibration (RMSEC) were 0.0197 for PSO-PLS and 0.0388 for W-PLS, respectively; the root mean standard errors of prediction (RMSEP) were 0.0193 for PSO-PLS and 0.0304 for W-PLS, respectively. The prediction results revealed that the prediction acidity values were closer to the chemical values for PSO-PLS than for W-PLS. Compared with W-PLS, PSO-PLS not only reduced wavelength variables and decreased calculation time, but also greatly improved the model performance and strengthened the prediction accuracy of the model. The results show that PSO-PLS can establish a good analysis model to accurately predict the acidity in apples and can be widely applied to rapidly analysis apple internal qualities in the field.

**Keywords**: Near-infrared; Particle swarm optimization; Apple; Acidity