蚁群和遗传算法优化花茶花青素近红外光谱预测模型的比较

李艳肖，黄晓玮，邹小波*，赵杰文，石吉勇，张小磊
（江苏大学 食品与生物工程学院，江苏 镇江 212013）

摘要：以建立花茶花青素含量的最优近红外光谱模型为目标，对比研究了蚁群算法（Ant Colony Optimization, ACO）和遗传算法（Genetic Algorithm, GA）优化近红外光谱谱区的效果。ACO-iPLS将全光谱划分为12个子区间时，优选出第1、9、10共3个子区间，所建的校正集和预测集相关系数分别为0.901 3和0.864 2；交互验证均方根误差（RMSECV）和预测均方根误差（RMSEP）分别为0.160 0 mg/g和0.202 0 mg/g；GA-iPLS将全光谱划分为15个子区间时，优选出第1、5共2个子区间，所建模型的校正集和预测集相关系数分别为0.906 3和0.879 3，交互验证均方根误差（RMSECV）和预测均方根误差（RMSEP）分别为0.156 0 mg/g和0.206 0 mg/g。研究结果表明：ACO-iPLS和GA-iPLS均可以有效选择近红外光谱特征波长，其中GA-iPLS模型的精度更高。

关键词：蚁群算法，遗传算法，区间偏最小二乘法，花茶，花青素，定量分析模型
中图分类号：O657·3　　**文献标识码：**A　　**文章编号：**1673—1689（2015）06—0575—09

Optimization of NIR Spectroscopy Based on Ant Colony Optimization and Genetic Algorithm for the Anthocyanin Content in Scented Tea

LI Yanxiao, HUANG Xiaowei, ZOU Xiaobo, ZHAO Jiewen, SHI Jiyong, ZHANG Xiaolei*
（College of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China）

Abstract: Optimization of Near infrared (NIR) spectroscopy for quantitative analysis of the anthocyanin content in scented tea was discussed by selecting the optimal spectra intervals from the whole NIR spectroscopy using two variable models: Ant colony optimization interval partial least squares (ACO-iPLS) and Genetic Algorithm interval partial least squares (GA-iPLS). The ACO-iPLS full-spectrum was split into 12 intervals. The optimal intervals selected were the 1st interval, 9th interval and 10th interval. The calibration and prediction correlation coefficient of ACO-iPLS model were 0.901 3 and 0.864 2, in which the root mean square error of cross validation (RMSECV) of 0.160 0 mg/g and the root mean square error of prediction (RMSEP) of 0.206 0 mg/g were achieved.

收稿日期：2014-06-22
基金项目：国家自然科学基金项目（60901079）；国家863计划项目（2011AA108007）；江苏省杰出青年基金项目（BK2013010）；江苏特聘教授基金项目（201205）
作者简介：李艳肖（1976-），女，河北石家庄人，工学硕士，实验师，主要从事农产品、食品检测。E-mail: li_yanxiao@163.com
***通信作者：**邹小波（1974-），男，湖南汨罗人，工学博士，教授，博士研究生导师，主要从事农产品、食品品质无损检测研究。E-mail: zou_xiaobo@ujs.edu.cn

As in the GA-iPLS model, the data set was split into 15 intervals for optimization where 1st and 5th intervals were selected. The calibration and prediction correlation coefficient of GA-iPLS model were 0.901 3 and 0.864 2, and the RMSECV and RMSEP of GA-iPLS models based on these intervals were 0.156 0 mg/g and 0.206 0 mg/g, respectively. The results showed that both ACO-iPLS and GA-iPLS models could efficiently select spectrum intervals for quantitative analysis of anthocyanin in scented tea. The optimal GA-iPLS model had better performance with higher accuracy.

Keywords: ant colony optimization, genetic algorithm, interval partial least squares, scented tea, anthocyanin, quantitative analysis model

近红外光谱法(near infrared spectroscopy, NIR)是一种快速无损的检测方法,随着计算机软件技术的发展,其在石油、医药、农产品等的检测方面显示出了巨大的潜力[1-5]。由于样本成分的复杂性和相干性,在对待测样品近红外光谱数据建立预测模型时,为了减少运算时间和剔除噪声过大的谱区,需要确定组分的特征谱区[6-8]。此外,优选特征谱区也具有一定实际运用价值,因为工程实际应用中的滤光片和LED光源都有一定的带宽,优选到的特征谱区可以为挑选合适的滤光片和LED光源提供参考。

国内外学者对近红外光谱特征波长选择方法的研究有很多,参考文献[9]中有详细介绍,每种方法均有各自的优劣[10]。例如,目前比较常用的光谱信息提取方法有区间偏最小二乘法(iPLS),该方法虽然能有效提取光谱中与特定组分最相关的谱区,但建模时光谱区间的挑选方法比较单一,往往只考察单个子区间或者少数几个子区间,使得挑选出来的子区间不能全面表征特征信息,模型难以达到最佳的预测效果。

近年来,基于仿生优化而发展起来的蚁群算法(Ant Colony Optimization, ACO)和遗传算法(Genetic Algorithm, GA)是特征变量筛选方法的研究热点,虽然都是基于群体智能发展起来的算法,但是两者的基本原理不相同。蚁群算法是由意大利学者M. Dorigo提出的一种模拟蚂蚁群体智能行为的仿生算法,它是一个增强型学习系统,具有良好的鲁棒性和分布式计算特性,但蚁群算法求解时间长,容易出现停滞现象[11];遗传算法是由John Holland教授提出的,以达尔文的生物进化论和孟德尔的遗传理论为基础,模拟自然界中生物遗传机制的仿生优化算法,该算法具有与问题域无关的全局搜索能力,且不易陷入局部最优,使用评价函数作为启发信息[12]。本文中将这两种算法分别与iPLS相结合,以建立最优预测花茶花青素含量的近红外光谱模型为目标,比较这两种算法对近红外光谱谱区的筛选效果。

1 蚁群区间偏最小二乘法(ACO-iPLS)和遗传区间偏最小二乘法(GA-iPLS)

1.1 ACO-iPLS基本原理

蚁群优化算法是模仿蚂蚁觅食方式的一种新的启发式算法。研究发现,蚂蚁在其觅食路上会留下一种被称为信息素(Pheromone)的物质,并且在搜寻的过程中能够感知出信息素的存在及强度,以此作为其选择路径的参考。蚂蚁在觅食过程中,通常朝着信息素强度大的方向运动,某一条路经过的蚂蚁越多信息素便越强,后来者选择该路径的概率也就越大。此外,信息素还会挥发,路径越长、时间越长,信息素挥发得愈多,信息素的强度就越小。信息素的累积和挥发的总和成为信息交流的媒体。相互协作的蚁群就是通过这种信息正反馈原理来完成最佳路径搜寻的[13-14]。

图1是ACO-iPLS的流程。假设将光谱分为m个区间,有k只蚂蚁进行优化。其算法简要介绍如下:

1)信息素含量初始化:每个区间拥有相同的信息素含量,$\tau_i(0)=\phi$ ($i=1,2,\cdots\cdots,m$)。

2)解的选择:每只蚂蚁基于概率函数选取特征区间,最简单有效的概率选择方法为

$$P_i(t)=\frac{\tau_i(t)}{\sum_{i=1}^{m}\tau_i(t)} \quad (1)$$

式(1)中,$\tau_i(t)$是区间i在时间t时所拥有的信息素含量。这一选择过程简要描述如下:①多个区间的选择概率和通过公式

$$accu(i)=accu(i-1)+P(i) \quad (2)$$

计算，式(2)中，accu(0)=0，当然，accu(m)=1(m是所有区间总数)；②产生一个随机(0, 1)之间的随机变量，如果这个随机变量在accu(i−1)和accu(i)之间，则第i个区间入选。这种选择方法广泛用于遗传算法中，显然，光谱区间拥有的信息素量越多就越容易被选到，可以通过权重函数来调整各个区间的选择概率。

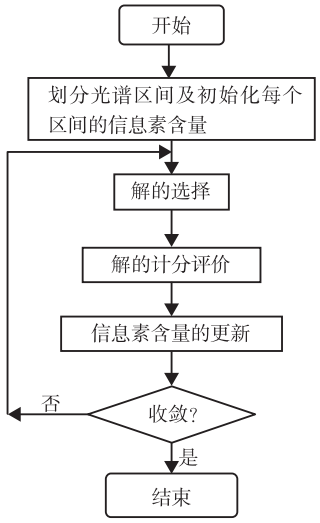

图1　ACO优选光谱区间的流程图

Fig.1　Flowchart of ACO oriented optimization process

3)解的计分评价：建立一个标准或客观函数来评价多得的解。本研究中基于已选区间上的iPLS模型预测精度作为每只蚂蚁所选区间的评价参数，即模型的预测均方根误差RMSEP的倒数作为评价函数。

4)信息素含量的更新：每个区间的信息素含量的更新通过公式(3)计算。

$$\tau_i(t+1)=r\cdot\tau_i(t)+\Delta\tau_i(t) \quad (3)$$

式(3)中，i为第i个光谱区间，r为区间信息素含量遗留率，对应信息素含量挥发率e (r=1−e)，r为(0, 1)之间的常数，$\Delta\tau_i(t)$为信息素含量的增长值，他与每只蚂蚁在所选区间上建立的iPLS模型精度成正比。即建立在该区间上的模型精度越高则信息素含量越多。在整个选择过程中，信息素含量的挥发率可以通过预先设定的挥发率e来获得。

重复以上2)—4)步骤中k个蚂蚁的“选择”、“计分评价”和“更新”这一迭代过程，最终通过一定数量的迭代后，理论上所有蚂蚁都会收敛到相同的区间变量上，从而得到最佳光谱区间。更多有关ACO的算法的原理请见参考文献[15-16]。

1.2　GA-iPLS的基本原理

本文中所用的遗传区间偏最小二乘波长筛选法是对Nϕrgaard提出的一种波长筛选法的改进和发展，该法主要用于筛选偏最小二乘建模的波长区域。其算法如下：

1.2.1　特征波谱区间入选编码　首先将整个花茶近红外光谱等分为s个区间，对这s个区间入选的问题，可用一含有s个0/1字符(基因)的字符串(染色体串)来表示每种区间组合。字符串0和1分别代表对应区间未被选中和选中，例如对8个区间的问题区间组合“00110101”，表示第3，4，6，8个区间被选中，其余则未被选中。

1.2.2　适应度函数的设计　采用PLS交互验证中因变量的预测值和实际值的相关系数(r)为适应度函数。具体实施方法为，对每个个体所选的区间进行数据重新组合，再用PLS交互验证得到相关系数(r)。相关系数(r)的计算公式为

$$r=\frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2\sum_{i=1}^{N}(y_i-\bar{y})^2}} \quad (4)$$

式(4)中，N为样品个数；$\bar{x}$为交互验证预测值的均值；$\bar{y}$为实际测量值的均值。

1.2.3　初始群体　本研究的初始种群由计算机随机产生的n个个体组成，而每个个体由s个字符组成。

1.2.4　遗传操作设计　选择算子采用最常用的选择方法——适应度比例方法，也称转轮法，每个个体的选择概率与其适应度成比例。交叉算子采用单点交叉方法，见表达式

$$0110|1 \xrightarrow{\text{交叉}} 01100$$
$$1100|0 \qquad\quad 11001 ,$$

参与交叉的个体概率为一个小于1的小数(如0.8)。

变异算子采用基本变异算子，即在某个个体(字符串)中随机挑选一个或多个基因(字符)进行变异，参与变异的个体概率也为一个小于1的小数，它通常比较小(如0.1)。

1.2.5　运算终止条件　本文中以遗传迭代次数达到设定的交互验证均方根误差(RMSECV)为收敛终止条件。

1.2.6　区间选取　本文中采用的方法为，在遗传迭

代后，具有最小RMSECV的区间组合中的所有入选区间为特征波谱区间[17-19]。

2 材料与方法

2.1 实验材料与仪器设备

从超市购买6种花茶（山茶花茶、洛神花花茶、月季花茶、玫瑰花茶、康乃馨花茶、勿忘我花茶），每种花茶分别用粉碎机粉碎，并过40目筛，然后按照四分法原则，随机称取2 g左右的粉末作为一个样本，每种花茶取10个平行样本，6种花茶共60个样本。

Antaris Ⅱ型傅里叶变换近红外光谱仪，美国赛默飞世尔公司制造；UV-1601型紫外分光光度计，日本岛津公司制造。

2.2 光谱采集及预处理

本实验中采用InGaAs检测器，波数范围10 000～4 000 cm^{-1}，扫描次数为32次；分辨率为8 cm^{-1}，波数间隔为3.853 6 cm^{-1}，每条光谱包含有1 557个变量。数据采集过程中，室内湿度保持基本不变，温度保持在25℃左右。由于花茶粉末为不透明颗粒，所以实验中采用积分球的漫反射采样方式，每个样本扫描一次后将样品池旋转120°，共扫描3次，取其平均光谱作为该样本的原始光谱，如图2所示。

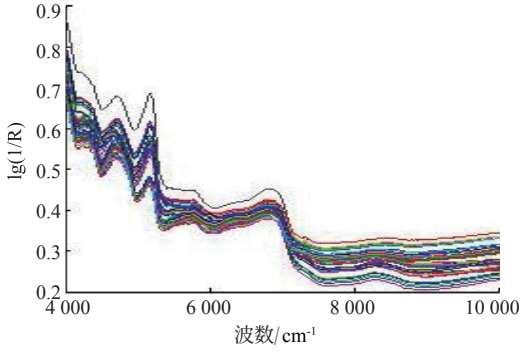

图2 花茶原始光谱图

Fig.2 Raw NIR spectra of scented tea samples

2.3 花青素含量的测定

准确称取2 g花茶粉末，加入100 mL含有1 mol/L柠檬酸的体积分数70%乙醇溶液，放入60℃水浴锅中浸提4 h后，经真空抽滤，60℃旋转蒸发10 min，取0.5 mL浓缩后的溶液置于2个50 mL容量瓶中，分别用pH值为1的盐酸缓冲液和pH为4.5的醋酸钠缓冲液定容至刻度线，分别静置50 min和80 min，用分光光度计在513 nm和730 nm下测定吸光度值，花青素质量分数[20-23]

$$w=((A/\varepsilon L)\times M_W\times D_f\times V/W_t)\times 100\% \quad (5)$$

式(5)中：

$$A=(A_{513}\ pH_{1.0}-A_{700}\ pH_{1.0})-(A_{513}\ pH_{4.5}-A_{700}\ pH_{4.5})$$

A为吸光度；D_f为稀释因子；ε为矢车菊花青素-3-葡萄糖苷的消光系数，26 900；M_W为矢车菊花青素-3-葡萄糖苷的相对分子质量，449.4；V为最终体积，mL；W_t为产品质量，mg；L为光程，1 cm。

3 结果与讨论

3.1 检测原始数据和光谱预处理

60个样本花茶中的花青素含量统计结果如表1所示，花青素含量范围为0.175 9~1.603 69 mg/g，每种花茶随机选择其中6个样本作为校正集，4个样本作为预测集，校正集共36个样本，预测集共24个样本。

表1 花茶样本花青素含量实测值数据统计

Table 1 Statistics of anthocyanin content for calibration and prediction set of scented tea

模型	样品数	最小值/(mg/g)	最大值/(mg/g)	平均值/(mg/g)	方差/(mg/g)
校正集	36	0.1 759	1.6 036	0.8 665	0.1 402
预测集	24	0.2 544	1.1 151	0.7 378	0.0 735

试验中，花茶样本颗粒的粒径大小和样本的密实度不可能完全一致，将会影响到光在固体颗粒内的漫反射。因而，需要对样本的原始光谱数据进行预处理，本实验中采用SNV预处理。SNV首先从原光谱中减去该条光谱的平均值，再除以标准偏差，主要用于消除由于样品颗粒大小不均匀和密实度不一样对光谱的影响[24]。预处理后的光谱如图3所示。

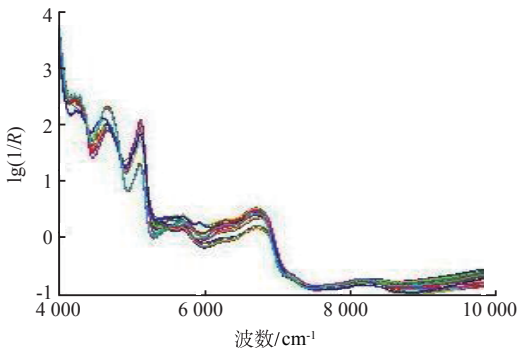

图3 经SNV预处理后的花茶光谱

Fig.3 NIR spectra of scented tea samples after SNV spectral

3.2 ACO-iPLS选择特征子区间

花茶粉末近红外光谱不仅包含了花青素信息，还包含了除花青素以外的组分信息。由于花青素在近红外光谱的多个波长处有吸收，且近红外光谱的谱峰较宽，故难以确定花青素特征子区间宽度。为了使ACO-iPLS准确定位包含花青素特征波长的子区间，需要对子区间划分总数进行优化。

ACO-iPLS通过选择光谱子区间宽度、蚂蚁个数、遗传代数、循环次数和变量数，选择最佳光谱区间进行建模。目标函数是判断种群中个体优劣和群体优化程度的标准，选择一个合适的目标函数可以加速算法收敛，提高计算精度。目标函数包括方差、标准差等。试验结果表明，选择均方根误差作为目标函数能够较快地收敛，因此选择均方根误差作为目标函数。

将预处理后的光谱数据划分为p个子区间，p的取值范围为10～30。当p取不同值时，采用ACO-iPLS选择的特征子区间，如表2所示。

表2 ACO-iPLS子区间优选结果

Table 2 Optimal spectra regions by ACO-iPLS method

区间总数	最佳ACO-iPLS			最佳GA-iPLS		
	入选区间	Rmsep/(mg/g)	入选波数点数	入选区间	Rmsep/(mg/g)	入选波数点数
10	4 000~4 590,4 594~5 184,9 345~9 935	0.207 0	462	4 000~4 597,4 601~5 199,9 407~1 000	0.206 0	467
11	4 000~4 524,4 528~5 053,9 284~9 808	0.217 0	411	4 000~4 543,5 643~6 187,6 190~6 734,8 917~9 457	0.209 0	567
12	4 000~4 466,7 764~8 231,8 235~8 701	0.202 0	366	4 000~4 497,5 504~6 001,7 509~8 007,9 010~9 503	0.230 0	519
13	4 000~4 428,4 864~5 292,7 455~7 872	0.208 0	336	4 000~4 459,5 851~6 310,6 314~6 773,9 087~9 542	0.209 0	479
14	4 000~4 389,7 540~7 930,7 934~8 323,8 327~8 717	0.220 0	408	4 000~4 428,5 742~6 148,6 152~6 576,8 721~9 145,9 149~9 573,9 577~1 000	0.207 0	667
15	4 000~4 385,7 895~8 281,8 285~8 670	0.210 0	303	4 000~4 397,5 604~6 001	0.206 0	208
16	4 000~4 351,7 548~7 899,8 258~8 609	0.208 0	276	4 000~4 374,5 512~5 886,8 883~9 253,9 631~1 000	0.210 0	390
17	4 000~4 343,7 471~7 814,7 818~8 161	0.230 0	270	4 000~4 351,5 419~5 770,5 774~6 125,6 129~6 480,8 601~8 948,9 303~9 650,9 654~1 000	0.213 0	641
18	4 000~4 320,6 915~7 236,7 239~7 560	0.209 0	252	4 000~4 331,4 335~4 667,4 671~5 002,5 677~6 009,7 351~7 679,7 683~8 001,9 010~9 338	0.240 0	606
19	4 000~4 308,4 937~5 245,7 436~7 745	0.215 0	243	4 000~4 312,4 632~4 945,4 948~5 261,5 897~6 210,6214~6 526,7 162~7 475,7 795~8 107,8 744~9 056,9 376~9 689,9 692~10 000	0.230 0	819
20	4 578~4 864,8 628~8 913,8 917~9 203	0.234 0	225	4 000~4 297,4 300~4 597,5 504~5 801,4 300~4 597,6 406~6 703,8 512~8 809,8 813~9 110,9 411~9 704,9 708~10 000	0.214 0	622
21	4 547~4 817,8 655~8 925,9 203~9 473	0.213 0	213	4 000~4 285,4 289~4 574,5 724~6 005,6 295~6 576,6 580~6 861	0.214 0	372
22	4 540~4 806,8 859~9 126,9 129~9 395	0.212 0	210	4 000~4 270,6 190~6 460,6 464~6 734,6 738~7 008,7 560~7 830,7 833~8 103,8 107~8 377,8 655~8 921,9 195~9 461,9 735~10 000	0.212 0	707
23	4 000~4 246,4 250~4 497,4 752~4 999,7 259~7 506	0.231 0	325	4 000~4 258,4 524~4 783,4 786~5 045,5 836~6 094,6 360~6 618,6 622~6 881,9 230~9 484,9 746~10 000	0.225 0	542
24	4 243~4 482,4 486~4 725,4 729~4 968,4 972~5 211,7 401~7 641,8 130~8 370,8 373~8 613,9 345~9 584,9 588~9 827	0.223 0	567	4 000~4 246,4 250~4 497,5 253~5 500,5 504~5 751,5 755~6 001,8 763~9 010,9 264~9 507,9 758~10 000	0.228 0	518
25	4 243~4 713,8 782~9 018,9 260~9 496,9 739~9 974	0.214 0	248	4 000~4 293,4 486~4 725,5 458~5 697,5 701~5 936,6 179~6 414,7 614~7 849,8 809~9 045,9 048~9 284	0.217 0	438
26	4 000~4 219,5 342~5 562,7 355~7 575,7 579~7 799,8 250~8 470,8 474~8 694,9 145~9 365,9 368~9 588,9 592~9 812	0.216 0	522	4 000~4 227,4 231~4 459,4 462~4 690,5 388~5 616,7 008~7 236,7 471~7 698,7 702~7 930,8 397~8 624,8 859~9 087,9 777~10 000	0.232 0	599

续表 2

区间总数	最佳ACO-iPLS			最佳GA-iPLS		
	入选区间	Rmsep/(mg/g)	入选波数点数	入选区间	Rmsep/(mg/g)	入选波数点数
27	4 216~4 428,4 864~5 076,7 671~7 884	0.232 0	168	4 000~4 219,5 342~5 562,5 566~5 785,6 013~6 233,6 460~6 680,7 803~8 022,8 026~8 242,8 246~8 462,8 466~8 682,8 906~9 122	0.226 0	576
28	4 000~4 208,4 212~4 220,7 606~7 814	0.214 0	165	4 000~4 212,4 216~4 428,4 648~4 860,5 728~5 940,6 160~6 372,6 376~6 588,7 023~7 236,7 455~7 668,7 671~7 880,7 884~8 092,8 944~9 153,9 368~9 577,9 581~9 789	0.210 0	722
29	4 200~4 397,4 802~4 999,7 409~7 606,8 211~8 408	0.218 0	208	4 000~4 204,4 208~4 412,5 874~6 079,6 082~6 287,6 291~6 495,6 915~7 120,7 124~7 328,7 540~7 745,8 778~8 979,8 983~9 183,9 392~9 592,9 800~10 000	0.251 0	644
30	4 590~4 783,5 377~5 569,9 114~9 307,9 507~9 700	0.205 0	204	4 000~4 196,4 200~4 397,5 604~5 801,5 805~6 001,6 005~6 202,6 406~6 603,7 409~7 606,7 810~8 007,8 211~8 408,8 813~9 010,9 014~9 210,9 214~9 411,9 415~9 608	0.225 0	

可以看出,当光谱划分为12(即p=12)时,对应的预测均方根误差(Rmesp)最小,选择窗口宽带为122 cm^{-1},每个光谱子区间对所建模型的权重系数是不相同的,如图4所示。

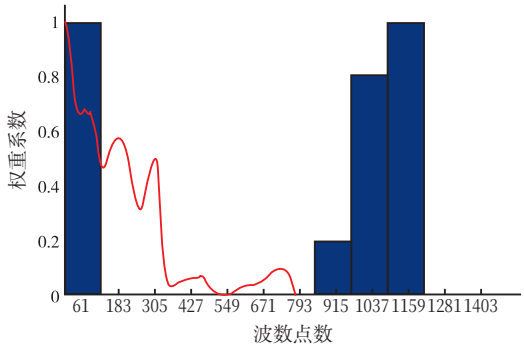

图4 ACO-iPLS光谱区间选择

Fig. 4 intervals selected by ACO-iPLS

选择权重系数≥0.3的子区间进行建模,满足上述条件的共有3个子区间:第1子区间(系数为1.00),第9子区间(系数为0.81),第10子区间(系数为1.00),这3个子区间对应的光谱范围分别为:4 000~4 466 cm^{-1},7 764~8 231 cm^{-1},8 235~8 701 cm^{-1}。在上述3个子区间的基础上,用iPLS建模,将所选光谱区间集合划分为25个子区间时,模型的预测精度和计算效率最高,校正集预测值和实测值之间的相关系数为0.901 3;预测集预测值和实测值之间的相关系数为0.864 2,其Rmsecv和Rmsep分别为0.160 0 mg/g和0.202 0 mg/g,如图5所示。

3.3 GA-iPLS谱区筛选模型优选特征区间

应用GA-iPLS对花茶的近红外光谱谱区进行筛选时,将全光谱分别划分为10、11、12、…、30个子区间,以考查不同数目的子区间划分对模型性能以及最佳波长区间的影响。GA-iPLS选择的特征子区间见上表2。

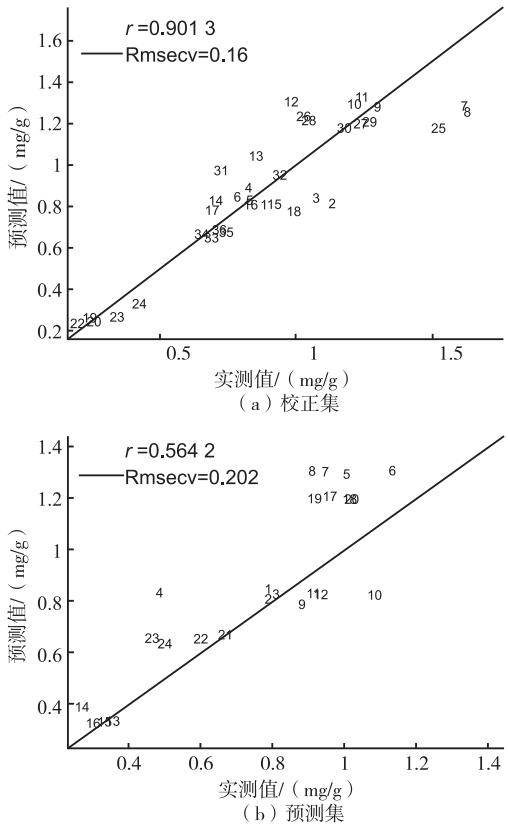

图5 ACO-iPLS最佳模型的预测值与实测值之间的关系

Fig.5 Reference measured versus NIR predicted by ACO-iPLS(a) calibration sets (b) prediction sets

将全光谱划分为15个子区间，主成分数为11，初始群体为40，交叉概率为0.95，变异概率为0.1；迭代次数为60次时，模型的预测精度和计算效率最高，校正集预测值和实测值之间的相关系数为0.906 3；预测集预测值和实测值之间的相关系数为0.879 3，其Rmsecv和Rmsep分别为0.156 0 mg/g和0.206 0 mg/g，如图6所示。

3.4 模型比较

为了对建模效果进行比较，分别对全光谱PLS、iPLS进行建模，结果如表3所示。

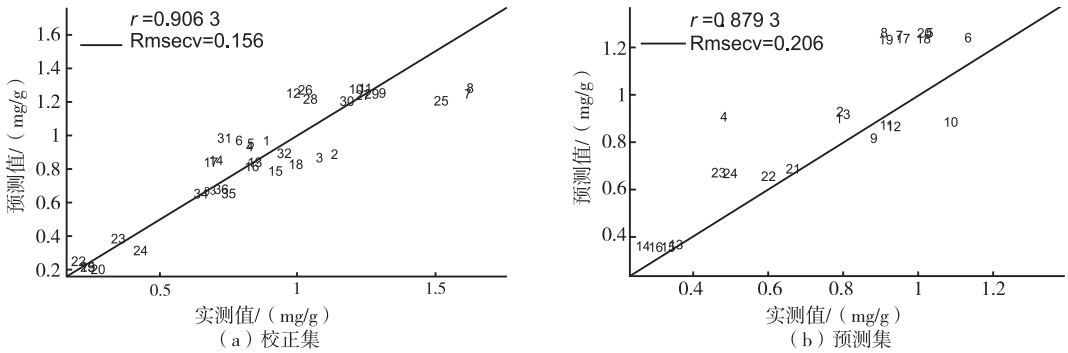

图6 GA-iPLS最佳模型样本的预测值与实测值之间的关系

Fig.6 Reference measured versus NIR predicted by GA-iPLS in prediction and calibration sets

表3 不同谱区筛选的模型的结果

Table 3 Results of different selecting wavelength regions models

模型	波数变量数	波数范围/cm^{-1}	校正集		预测集	
			R	Rmsecv(mg/g)	R	Rmsep(mg/g)
全光谱PLS	1 557	4 000~10 000	0.793 4	0.225 0	0.788 5	0.231 0
iPLS	86	4 424~4 632	0.893 3	0.167 0	0.875 2	0.249 0
ACO-iPLS	366	4 000~4 466，7 764~8 231，8 235~8 701	0.901 3	0.160 0	0.864 2	0.202 0
GA-iPLS	208	4 000~4 397，5 604~6 001	0.906 3	0.156 0	0.879 3	0.206 0

从表3可以看出，ACO-iPLS和GA-iPLS模型精度都比全光谱PLS和iPLS模型精度高，其中GA-iPLS最高。而GA-iPLS模型和ACO-iPLS模型需要的光谱区间所含的波数点分别为208和366，只有全光谱（波数点数为1 557个）的约1/7和1/4；iPLS建模虽只需一个光谱区间，但是模型的精度和预测效果较差。

对比结果表明，ACO-iPLS和GA-iPLS均可以大大简化模型，同时还能提高模型的预测精度和鲁棒性，且两种优化算法都包含有4 000～4 400 cm^{-1}（2 270～2 500 nm）这个波段。

光谱分析表明，该近红外波段频率的4倍对应的是花青素在可见光波段的特征吸收波段（560～6 200 nm）。

4 结 语

将ACO和GA分别与iPLS相结合，以建立最优花茶花青素含量的近红外光谱预测模型为目标，对比研究这两种算法对近红外光谱区筛选的结果，ACO-iPLS模型对应的校正集和预测集相关系数分别为0.901 3和0.864 2，Rmsecv和Rmsep分别为0.160 0 mg/g和0.202 0 mg/g；GA-iPLS模型对应的校正集和预测集相关系数分别为0.906 3和0.879 3，Rmsecv和Rmsep各为0.156 0 mg/g和0.206 0 mg/g，均优于全光谱PLS和iPLS模型。研究结果充分表明，将ACO-iPLS和GA-iPLS对花茶花青素含量进行建模，简化了模型复杂度，提高了模型的预测精度和计算效率，其中GA-iPLS模型的精度更高。

参考文献:

[1] 石吉勇, 邹小波, 赵杰文, 等. 基于近红外光谱的设施栽培水果黄瓜磷元素亏缺初期快速诊断[J]. 光谱学与光谱分析, 2011, 31(12): 3264-3268.
SHI Jiyong, ZOU Xiaobo, ZHAO Jiewen et al. Rapid diagnostics of early phosphorus deficiency in mini-cucumber plants under protected cultivation by near infrared spectroscopy[J]. **Spectroscopy and Spectral Analysis**, 2011, 31 (12): 3264- 3268. (in Chinese)

[2] 刘燕德, 陈兴苗, 孙旭东. 可见/近红外漫反射光谱无损检测南丰蜜桔维生素C的研究[J]. 光谱学与光谱分析, 2008, 28(10): 2318-2324.
LIU Yande, CHEN Xingmiao, SUN Xudong. Nondestructive measurement of vitamin c in nanfeng tangerine by visible/near-infrared diffuse reflectance spectroscopy[J]. **Spectroscopy and Spectral Analysis**, 2008, 28(10): 2318-2324.(in Chinese)

[3] 汤旭光, 宋开山, 刘殿伟, 等. 基于可见/近红外反射光谱的大豆叶绿素含量估算方法比较[J]. 光谱学与光谱分析, 2011, 31(2): 371-374.
TANG Xuguang, SONG Kaishan, LIU Dianwei, et al. Comparison of methods for estimating soybean chlorophyll content based on visual / near infrared reflection spectra[J]. **Spectroscopy and Spectral Analysis**, 2011, 31(2): 371-374.(in Chinese)

[4] 李水芳, 张欣, 单杨, 等. 近红外光谱检测蜂蜜中可溶性固形物含量和水分的应用研究[J]. 光谱学与光谱分析, 2010, 30(9): 2377-2380.
LI Shuifang, ZHANG Xin, SHANG Yang, et al. Prediction Analysis of soluble solids content and moisture in honey by near infrared spectroscopy[J]. **Spectroscopy and Spectral Analysis**, 2010, 30(9): 2377-2380.(in Chinese)

[5] 芮玉奎, 辛术贞, 李军会. 应用近红外光谱技术测试温室黄瓜叶片全氮含量[J]. 光谱学与光谱分析, 2011, 31(8): 2114-2116.
RUI Yukui, XIN Shuzhen, LI Junhui. Application of NIRS to detecting total n of cucumber leaves growing in greenhouse[J]. **Spectroscopy and Spectral Analysis**, 2011, 31(8): 2114-2116.(in Chinese)

[6] 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): 458-466.
SHI Jiyong, ZOU Xiaobo, ZHAO Jiewen et al. Selection of wavelength for strawberry NIR spectroscopy based on BIPLS combined with SAA[J]. **j. Infrared Millim Waves**, 2011, 30(5): 458-466.(in Chinese)

[7] 石吉勇, 邹小波, 赵杰文 等. 一种近红外光谱特征子区间选择新算法[J]. 光谱学与光谱分析, 2010, 54(12): 3119-3125.
SHI Jiyong, ZOU Xiaobo, ZHAO Jiewen et al. A new method of characteristic wavelength sub-range selection of near infrared spectroscopy[J]. **Spectroscopy and Spectral Analysis**, 2010, 54(12): 3119-3125.(in Chinese)

[8] Shi Jiyong, Zou Xiaobo, Zhao Jiewen,et al. Diagnostics of nitrogen deficiency in mini-cucumber plant by near infrared reflectance spectroscopy[J]. **African Journal of Biotechnology**, 2011(10): 19687-19692.

[9] Zou Xiaobo, Zhao Jiewen, Povey Malcolm J W,et al. Variables selection methods in near- infrared spectroscopy[J]. **Analytica Chimica Acta**, 2010, 667(1-2): 14-32.

[10] Zou Xiaobo, Zhao Jiewen, Huang Xingyi,et al. Use of FT-NIR spectrometry in non-invasivemeasurements of solublesolid contents (SSC) of 'Fuji' apple based on different PLS models[J]. **Chemometrics And Intelligent Laboratory Systems**, 2007, 87(1): 43-51.

[11] 刘甲林. 基于改进蚁群算法的油品调和配方优化研究[D]. 大连:大连理工大学,2011.

[12] 栾东磊. 近红光谱分析技术在几种水产品中的应用研究[D]. 大连:大连理工大学,2009.

[13] 陈永明, 林萍, 何勇. 基于遗传算法的近红外光谱橄榄油产地鉴别方法研究[J]. 光谱学与光谱分析, 2009, 29(3): 671-674.
CHEN Yongming, LIN Ping, HE Yong. Study on discrimination of producing area of olive oil using near infrared spectra based on genetic algorithms[J]. **Spectroscopy and Spectral Analysis**, 2009, 29(3): 671-674.(in Chinese)

[14] Allegrini Franco, Olivieri Alejandro C. A new and efficient variable selection algorithm based on ant colony optimization. Applications to near infrared spectroscopy/partial least-squares analysis[J]. **Analytica Chimica Acta**, 2011, 699(1): 18-25.

[15] 郭亮, 吉海彦. 蚁群算法在近红外光谱定量分析中的应用研究[J]. 光谱学与光谱分析, 2007, 27(9): 1703-1705.
GUO Liang, JI Haiyan. Application study of ant colony algorithm in near infrared spectroscopy quantitative analysis[J]. **Spectroscopy and Spectral Analysis**, 2007, 27(9): 1703-1705.(in Chinese)

[16] 朱峰, 陈莉. 蚁群与遗传算法融合的聚类算法研究[J]. 西北大学学报:自然科学版, 2009, 39(5): 745-749.
ZHU Feng, CHEN Li. Research on clustering algorithm based on fusion of ant colony and genetic algorithm[J]. **Journal of**

Northwest University (**Natural Science Edition**), 2009, 39(5): 745-749.(in Chinese)

[17] 李艳肖, 邹小波, 董英. 用遗传区间偏最小二乘法建立苹果糖度近红外光谱模型[J]. 光谱学与光谱分析, 2007, 27(10): 2001-2004.
LI Yanxiao, ZOU Xiaobo, DONG Ying. Near infrared determination of sugar content in apples based on ga-ipls[J]. **Spectroscopy and Spectral Analysis**, 2007, 27(10): 2001-2004.(in Chinese)

[18] 屠振华, 籍保平, 孟超英,等. 基于遗传算法和间隔偏最小二乘的苹果硬度特征波长分析研究[J]. 光谱学与光谱分析, 2009, 29(10): 2760-2764.
TU Zhenhua, JI Baoping, MENG Chaoying, et al. Analysis of nir characteristic wavelengths for apple flesh firmness based on ga and ipls[J]. **Spectroscopy and Spectral Analysis**, 2009, 29(10): 2760-2764.(in Chinese)

[19] 王加华, 韩东海. 基于遗传算法的苹果糖度近红外光谱分析[J]. 光谱学与光谱分析, 2008, 28(10): 2308-2311.
WANG Jiahua, HAN Donghai. Analysis of near infrared spectra of apple ssc by genetic algorithm optimization[J]. **Spectroscopy and Spectral Analysis**, 2008, 28(10): 2308-2311.(in Chinese)

[20] 朱毛毛.桑椹红色素的提取纯化及其抗氧化活性和稳定性研究[D].镇江:江苏大学,2009.

[21] Patil Ganapathi, Madhusudhan M C, Babu B Ravindra ,et al. Extraction, dealcoholization and concentration of anthocyanin from red radish[J]. **Chemical Engineering and Processing: Process Intensification**, 2009, 48(1): 364-369.

[22] Sarma Annamraju D, Sreelakshmi Yellamraju, Sharma Rameshwar. Antioxidant ability of anthocyanins against ascorbic acid oxidation[J]. **Phytochemistry**, 1997, 45(4): 671-674.

[23] 陈健, 孙爱东, 高雪娟,等. 蓝莓花青素的提取及抗氧化性的研究[J]. 北京林业大学学报, 2011, 33(2): 126-129.
SU Jian, SUN Aidong, GAO Xuejuan, et al. Extraction and antioxidation of anthocyanins from blueberry[J]. **Journal of Beijing Forestry University**, 2011, 33(2): 126-129.(in Chinese)

[24] Chen Quansheng, Jiang Pei, Zhao Jiewen. Measurement of total flavonen content in snow lotus (Saussurea involucrate) using near infrared spectroscopy combined with interval PLS and genetic algorithm[J]. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, 2010, 76(1): 50-55.

会　议　信　息

会议名称(中文):第八届国际生物技术与农业峰会

会议名称(英文): The 8th International Biotechnology and Agriculture Summit (BAS 2015)
开始日期: 2015-08-26
结束日期: 2015-08-28
所在城市:北京市　海淀区
具体地点:北京海淀永泰福朋喜来登酒店
主办单位:北京市科学技术委员会
承办单位:北京生物技术和新医药产业中心、加州大学戴维斯分校动物医学学院
议题: Internationalization of Food Safety ;Implementation of Genomics in Food Safety & Security;Crop security & Food Animals ; Poster Presentation and Workshops;Food security & Public Health; Food Safety and Specific Threats;Tools for the 21st Century;Tools for the 21st Century