王胜鹏,郑鹏程,桂安辉,等.基于近红外光谱技术的远安黄茶品质快速无损检测方法[J].华中农业大学学报,2022,41(1):238-245.
DOI:10.13300/j.cnki.hnlkxb.2022.01.024

基于近红外光谱技术的远安黄茶品质快速无损检测方法

王胜鹏[1],郑鹏程[1],桂安辉[1],滕靖[1],刘盼盼[1],叶飞[1],高士伟[1],马梦君[2],刘小英[3]

1.湖北省农业科学院果树茶叶研究所,武汉 430064;2.咸宁市农业科学院,咸宁 437100;
3.恩施花枝山生态农业股份有限公司,恩施 445000

摘要　应用近红外光谱(near infrared spectroscopy,NIRS)技术结合多种算法对远安黄茶品质开展快速无损评价。首先通过扫描获得远安黄茶 90 个样品的近红外光谱,再利用 11 种不同方法对光谱进行预处理,剔除部分噪声信息,然后应用反向区间偏最小二乘法(backward interval partial least squares,Bi-PLS)筛选反映样品品质的特征光谱区间,应用遗传算法(genetic algorithm,GA)精准提取特征光谱波长,建立了 5 种黄茶品质预测模型,最后对光谱官能团信息进行解析。结果表明,最佳光谱预处理方法为多元散射校正,Bi-PLS 筛选出的特征光谱区间主要为 9 003.2～7 497.9 cm^{-1}、6 101.7～5 449.8 cm^{-1} 和 4 601.3～4 246.5 cm^{-1},GA 筛选出 75 个特征光谱波长,建立的 Bi-GA-PLS 组合模型具有最佳的稳健性,可准确地预测远安黄茶样品外部品质分数(R^2=0.951,RMSEP=1.57,RPD=5.27),初步实现了远安黄茶品质的快速、准确预测。光谱信息解析结果显示,45 个光谱波长反映—CH_x、C=O 和—NH_x 官能团信息,代表单糖、咖啡碱、茶氨酸和游离蛋白质等内含成分物质,30 个光谱波长反映 O—H、酰胺键以及 C—H 和 C—C 伸缩的组合频信息,代表木质素、淀粉、纤维素等多糖内含成分物质。

关键词　远安黄茶;品质评价;无损检测;近红外光谱;反向区间偏最小二乘法;遗传算法;人工神经网络
中图分类号　TS 272　　**文献标识码**　A　　**文章编号**　1000-2421(2022)01-0238-08

黄茶是我国六大茶类之一,湖北则以远安黄茶历史最为悠久。茶鲜叶经摊放、杀青、闷黄和烘干后加工为远安黄茶,其具有外形卷曲紧结带钩、色泽金黄、香气清香持久、滋味醇厚回甘、汤色杏黄明亮和叶底嫩黄匀整等特点[1-2],深受消费者喜爱。因此,对远安黄茶品质开展评价就显得非常重要。通常应用感官审评方法对农作物品质开展评价[3-4],在茶叶研究领域已实现对绿茶[5]、红茶[6]和乌龙茶[7]等茶类的品质评价。传统感官评价方法虽然经典,但专业性较强,评审结果易受审评员嗜好差异等因素影响,主观性较强。化学检测方法通过测定内含成分含量来评价茶品质,结果较客观、准确,但测定前需先粉碎样品,且测定过程费时、费力,还需使用大量的化学试剂,可能会造成环境污染,不利于茶品质的快速检测[8]。因此,非常有必要建立一种快速无损、科学客观的远安黄茶品质评价方法。

近红外光谱(near infrared spectroscopy,NIRS)技术是一种绿色的分析技术[9],主要反映样品内部成分 X—H 化学键信息,通过建立某一化学成分含量的预测模型从而实现样品的快速、无损检测。NIRS 技术目前已在农业、饲料和医药等[10-12]行业得到广泛应用。此外,NIRS 技术还应用于对茶叶中茶多酚、游离氨基酸、含水量、咖啡碱等含量的快速预测和茶鲜叶质量的快速评估[13]以及茶品质的评价[14-15]等方面。

目前,有关黄茶的研究主要集中在黄茶适制品种的筛选[16]、闷黄过程中品质成分的变化[17]、闷黄和加工工艺的优化[18]以及黄茶生理保健功效[19]等方面,还较少有应用近红外光谱技术在远安黄茶品质成分分析方面开展研究的报道。本研究基于近红外光谱技术,分别结合偏最小二乘法、反向区间偏最小二乘法(backward interval partial least squares,Bi-PLS)、遗传算法(genetic algorithm,GA)和人工神经网络方法(artificial neural network,ANN)筛选最佳光谱预处理方法和提取特征光谱数据点,建立远安黄茶品质预测模型,并尝试解析光谱官能团

收稿日期:2021-07-21
基金项目:国家茶叶产业技术体系项目(CARS-19);中央引导地方科技发展专项(2018ZYYD009);国家自然科学基金项目(31400586);湖北省农业科技创新中心项目(2016-620-000-001-032)
王胜鹏,E-mail:wwsspp0426@163.com.
通信作者:高士伟,E-mail:gsw0609@126.com

信息，以期为远安黄茶品质的快速无损评价提供新的思路。

1　材料与方法

1.1　试验材料

采集湖北省宜昌市远安县嫘祖黄茶有限公司基地的标准一芽二叶安吉白茶鲜叶，经杀青、闷黄等工序加工，获得 90 个远安黄茶样品，样品加工时间为 2016 年 4 月至 5 月。依据样品不同品质，将样品按照 4∶1 比例分为建模样品（72 个）和外部测试样品（18 个）。72 个建模样品用于建立预测模型，以品质分数为依据将样品按照 2∶1 比例划分为校正集（48 个样品）和验证集（24 个样品），其中验证集样品用于检验模型的稳健性。18 个测试样品用于检验模型的实际预测效果。

1.2　方　法

1）感官审评。根据 GB/T 23776 — 2018《茶叶感官审评方法》标准方法对远安黄茶品质进行评价。5 位感官审评专家采用密码评审方式对样品品质进行打分，满分为 100 分，品质越好，分数越高。以平均值作为该样品的最终品质得分。

2）近红外光谱采集。应用 Antaris Ⅱ型傅里叶变换近红外光谱仪，采用漫反射方式扫描获得样品的近红外光谱，仪器光谱范围为 4 000～10 000 cm^{-1}，分辨率为 8 cm^{-1}，检测器为 InGaAs。在扫描样品前，为使仪器达到最佳性能，保证状态稳定，通常先将仪器预热 1 h 后再扫描光谱，以镀金为光谱扫描背景。扫描时，将 10 .0 g 样品装入与仪器配套的旋转杯中，用压样器充分压实样品，保证光谱无法穿透样品，确保获得样品的全部光谱信息。每个样品重复装样 3 次扫描得到 3 条光谱，每条光谱扫描 64 次，取 3 条光谱的平均值作为该样品的最终光谱值。

3）光谱数据分析。应用 TQ Analyst 9.4.45 软件将每条近红外光谱转化为 1 557 对数据点保存于 Excel 表中，应用 OPUS 7.0 软件和 Matlab 2012a 软件对光谱数据进行预处理和建立预测模型。光谱扫描过程中往往会包含一些与样品性质无关的因素带来的干扰，如样品的状态、光的散射及仪器响应等的影响，导致光谱基线漂移和产生噪声信息。为有效去除光谱中夹杂的大量噪声信息，提高光谱的信噪比，分别比较标准变量变换（standard normal variate，SNV）、多元散射校正（multiple scatter correction，MSC）、消除常数偏移量（eliminate constant offset，ECO）、矢量归一化（vector normalization，VN）、减去一条直线（subtract straight line，SSL）、一阶导数（first derivative，FD）和二阶导数（second derivative，SD）等光谱预处理方法的建模效果，得到最佳预处理方法。

应用反向区间偏最小二乘法[20]（backward interval partial least squares，Bi-PLS）筛选反映远安黄茶品质的特征光谱区间。在 PLS 基础上，Bi-PLS 将全部光谱数据点等划分为 20～24 个光谱子区间，然后联合其中的 2～4 个光谱区间建立预测模型，当交互验证均方根误差（root mean square error of cross validation，RMSECV）最小时，此时得到的光谱区间即为筛选的最佳光谱区间。RMSECV 计算公式如下：

$$\mathrm{RMSECV}=\sqrt{\frac{\sum_{i=1}^{n}(y'_i-y_i)^2}{n}} \tag{1}$$

式（1）中：n 为校正集样品数；y_i 为样品 i 实测值；y_i' 为校正集样品 i 预测值。

应用遗传算法[21-22]（genetic algorithm，GA）对上述得到的最佳光谱区间进一步分析，剔除其中无用的光谱信息，并用特征光谱数据点建立 PLS 预测模型，所得结果用校正集决定系数（R_c^2）、预测集决定系数（R_p^2）、RMSECV、预测均方根均误差（root mean square error of prediction，RMSEP）表示。当 R^2 越大、RMSEP 越小时，表示模型预测效果越好。

RMSEP 计算公式如下：

$$\mathrm{RMSEP}=\sqrt{\frac{\sum_{i=1}^{n}(y_i-y'_i)^2}{n}} \tag{2}$$

式（2）中：n 为验证集样品数；y_i 为样品 i 实测值；y_i' 为验证集样品 i 预测值。

2　结果与分析

2.1　远安黄茶感官品质分析

从表 1 可知，远安黄茶全部样品的感官品质分数为 78.40～92.05，平均值为 88.96，标准偏差为 2.55；而以感官分数为分类依据，校正集样品感官分数范围为 78.40～92.05，平均值为 89.05，标准偏差为 2.61；验证集样品感官评分范围为 84.95～91.75，平均值为 89.11，标准偏差为 1.83。验证集样品品质分数处于校正集样品范围内，为建立一个稳健的预测模型提供了良好的前提条件。

表1 远安黄茶感官品质分数统计

Table 1 Statistics of sensory quality of Yuan'an yellow tea

数据统计 Data statistics	最大值 Maximum	最小值 Minimum	平均值 Average	标准偏差 Standard deviation
全部样品 All samples	92.05	78.40	88.96	2.55
校正集 Calibration set	92.05	78.40	89.05	2.61
验证集 Prediction set	91.75	84.95	89.11	1.83

2.2 光谱预处理方法筛选

从图1可以看出，不同品质远安黄茶近红外光谱整体变化趋势是一致的，随着黄茶品质的逐渐升高，光谱吸光度也逐渐增大，但在9 800～10 000 cm^{-1}和4 000～4 200 cm^{-1}这2个区间段存在较多的噪声信息。因此，在建模时应舍弃这2个波数区间。在6 900 cm^{-1}和5 100 cm^{-1}附近是游离水的—OH吸收峰，可能会影响品质模型的预测精度。此外，尝试应用多种预处理方法对样品光谱进行预处理，并用PLS方法建立品质分数预测模型，得出较佳的光谱预处理方法。

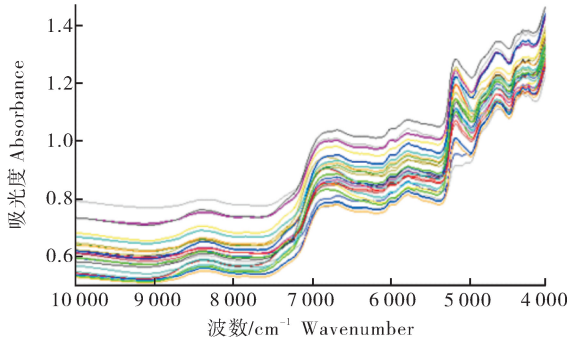

图1 远安黄茶样品近红外光谱

Fig.1 Near infrared spectroscopy of Yuan'an yellow tea

由表2可知，不同的光谱预处理方法建立的全波长PLS模型结果间差异较大。应用二阶导数预处理方法建立的PLS模型结果最差(R_c^2=0.614，RMSECV=2.75)，无光谱预处理建立的PLS模型结果相对较差(R_c^2=0.705，RMSECV=2.35)，应用多元散射校正预处理方法建立的PLS模型结果最佳(R_c^2=0.793，RMSECV=1.95)，RMSECV较最差的二阶导数模型降低了29.1%，可见对原始光谱进行有效去噪处理，可以提高模型预测效果，但并不是所有的预处理方法都会起到较好的效果。

2.3 Bi-PLS模型建立

应用多元散射校正预处理方法对光谱进行去噪处理后建立的PLS模型结果较差，还不能很好地预测远安黄茶品质。为此，尝试应用Bi-PLS方法进一步筛选与远安黄茶品质密切相关的光谱区间并建立预测模型(表3)。由表3可知，应用Bi-PLS方法筛选特征光谱区间(9 003.2～7 497.9 cm^{-1}、6 101.7～5 449.8 cm^{-1}、4 601.3～4 246.5 cm^{-1})后建立的预测模型结果均比PLS模型结果有所提高，最佳Bi-PLS模型较最佳PLS模型R_c^2提高了7.57%，最佳Bi-PLS模型较最佳PLS模型RMSECV下降了15.83%。可见，应用Bi-PLS方法不仅减少了建模的

表2 不同光谱预处理方法的处理结果

Table 2 Comparison of spectral pretreatment methods

序号 No.	预处理方法 Pretreatment methods	维数 Dimension	R_c^2	RMSECV
1	MSC	6	0.793	1.95
2	SNV	8	0.767	2.03
3	ECO	9	0.735	2.20
4	无 None	6	0.705	2.35
5	FD+MSC	5	0.688	2.46
6	VN	8	0.687	2.50
7	FD	7	0.659	2.62
8	FD+SNV	9	0.645	2.72
9	FD+SSL	5	0.634	2.75
10	SSL	1	0.625	2.73
11	SD	1	0.614	2.75

注 Note：MSC：多元散射校正 Multiple scatter correction；SNV：标准变量变换 Standard normal variate；ECO：消除常数偏移量 Eliminate constant offset；FD：一阶导数 First derivative；VN：矢量归一化 Vector normalization；SSL：减去一条直线 Subtract straight line；SD：二阶导数 Second derivative.

表3 远安黄茶品质Bi-PLS模型预测结果

Table 3 Results of Bi-PLS models for Yuan'an yellow tea

序号 No.	光谱区间/cm^{-1} Spectral regions	R_c^2	RMSECV
1	9 003.2～7 497.9；6 101.7～5 449.8；4 601.3～4 246.5	0.853	1.64
2	7 501.7～6 097.8；4 601.3～4 246.5	0.824	1.66
3	9 403.2～5 449.8；4 601.3～4 246.5	0.817	1.69
4	4 601.3～4 246.5	0.771	1.74
5	7 501.7～6 097.8；5 453.7～4 597.5	0.788	1.79
6	9 403.2～6 097.8；5 453.7～4 246.5	0.754	1.80
7	9 403.2～6 097.8；4 601.3～4 246.5	0.712	1.87
8	9 403.2～6 097.8	0.707	1.89
9	6 101.7～5 449.8；4 601.3～4 246.5	0.698	1.93

光谱数据量(650 个光谱数据点，占全部光谱数据点比例为 41.80%)，还提高了品质模型预测效果，但是对最佳模型效果的提升有限。

2.4　Bi-GA-PLS 模型建立

虽然 Bi-PLS 方法可以筛选特征光谱区间，但在光谱区间中还会存在较多与品质无关的光谱数据点，因此，非常有必要进一步精准提取出与品质密切相关的光谱数据点。用遗传算法对 9 003.2～7 497.9 cm^{-1}、6 101.7～5 449.8 cm^{-1} 和 4 601.3～4 246.5 cm^{-1} 3 个特征波段内的 650 个数据点进一步筛选。遗传迭代的参数设置为：初始群体 48，交叉概率0.8，变异概率 0.01，遗传迭代 100 次。通过观察遗传迭代后不同变量数的频率变化(图 2)和与 RMSECV 的变化关系(图 3)，当 RMSECV 最小时，此时参与建模的变量即为反映远安黄茶品质的最佳 NIRS 数据点。

从图 2 和图 3 可以看出，应用 GA 优化特征光谱区间的 650 个波数点，每个数据点的使用频率代表其对建模的重要性程度，其中绝大多数数据点的使用频率小于 5 次，表明与品质无关的数据点较多。当最佳建模数据点为 75 个时，Bi-GA-PLS 校正集模型 RMSECV 最小为 1.521，此时筛选得到的数据点即为最佳的反映远安黄茶品质的 NIRS 数据点。75 个光谱数据点的使用频率和与之对应的波数值见图 4。从图 4 可以看出，光谱数据点最大使用频率为 22 次，最小使用频率为 7 次，使用频率大于或等于 11 的光谱数据点共 14 个，占全部最佳数据点的比例为 18.67%。依据数据点的使用频率可以保证将最佳光谱数据点从大量的冗余信息中提取出来，有利于改善模型预测效果。在所有最佳光谱数据点中，最大光谱数据点为 8 990.5 cm^{-1}，最小光谱数据点为4 258.1 cm^{-1}。其中，在 9 003.2～7 497.9

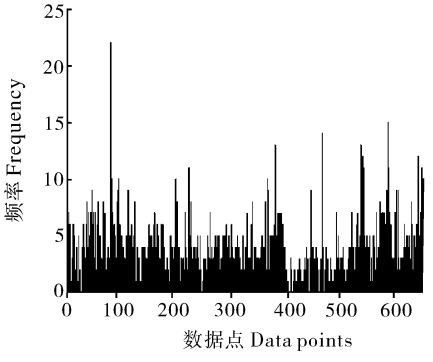

图 2　光谱数据点频率变化

Fig.2　Frequency variation of spectral data points

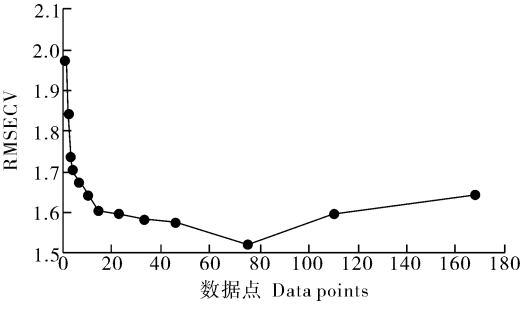

图 3　RMSECV 与建模最佳数据点关系

Fig.3　Relationship between RMSECV and the best data points of model

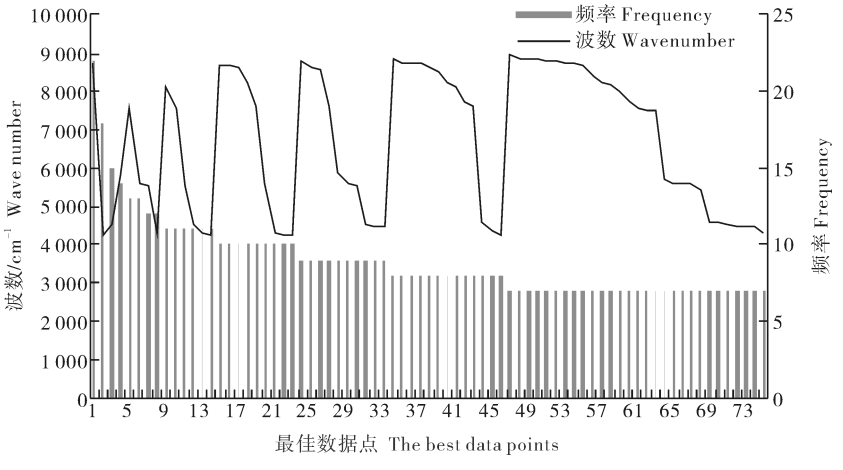

图 4　最佳光谱数据点的频率与波数

Fig.4　Frequencies and wavenumbers of the data points

cm^{-1}波段中共有390个数据点，有40个数据点被提取为最佳数据点；在6 101.7～5 449.8 cm^{-1}波段中共有168个数据点，有13个数据点被提取为最佳数据点；在4 601.3～4 246.5 cm^{-1}波段中共有92个数据点，有22个数据点被提取为最佳数据点。

2.5 Bi-GA-BP-ANN模型建立及模型优势比较

以遗传算法筛选的75个最佳光谱数据点为输入值，以远安黄茶品质为输出值，应用反向传播人工神经网络方法建立远安黄茶品质的近红外光谱Bi-GA-BP-ANN预测模型。在建立人工神经网络模型过程中，设置学习速率为0.1，传递函数为linear[－1，1]，模型所得结果见表4。建立远安黄茶品质分数全波长PLS模型、Bi-PLS模型、Bi-GA-PLS模型、全波长GA-PLS模型和Bi-GA-ANN模型（表4），并分别用验证集样品和外部样品对上述5种模型的稳健性进行检验，比较不同方法间建立模型的优越性。

从表4可以看出，在上述5种NIRS预测模型中，以Bi-GA-PLS组合模型建立的结果最佳（R_p^2＝0.964，RMSEP＝1.54，RPD＝5.27），以去噪处理后应用全部波长建立的PLS模型结果最差（R_p^2＝0.732，RMSEP＝1.98，RPD＝1.93）。GA-PLS模型预测结果稍优于Bi-PLS模型结果，这可能是由于GA筛选得到的特征光谱数据点中含有的噪声信息要少于应用Bi算法得到的特征光谱区间中的噪声信息。Bi-GA-ANN模型的结果稍差于Bi-GA-PLS模型结果，虽然ANN方法具有较好的非线性拟合能力，但是由于其内部复杂的信息传递过程[23]，反而不利于本研究建立远安黄茶品质预测模型。因此，本研究应用近红外光谱技术结合Bi-GA-PLS组合算法实现了远安黄茶品质的快速、无损、准确预测。

表4 5种模型预测结果优势比较

Table 4 Comparison of prediction results of five models

模型 Model	校正集 Correction set		验证集 Validation set		检验集 Test set		RPD
	R_c^2	RMSECV	R_p^2	RMSEP	R^2	RMSEP	
PLS	0.797	1.94	0.732	1.98	0.670	2.30	1.93
Bi-PLS	0.855	1.65	0.837	1.70	0.807	1.87	2.48
Bi-GA-PLS	0.971	1.52	0.964	1.54	0.951	1.57	5.27
GA-PLS	0.866	1.63	0.850	1.67	0.814	1.84	2.58
Bi-GA-BP-ANN	0.914	1.56	0.906	1.59	0.893	1.59	3.26

3 讨 论

3.1 特征光谱区间的筛选

每条样品NIRS都由1 557对数据点组成，数据量巨大，如果应用全部光谱数据建立模型，不仅造成建模时间较长，而且由于冗余信息的存在，致使模型欠稳健，不利于后期维护和使用。因此，非常有必要选取与样品密切相关的NIRS信息，而仅依靠去除样品中的噪声信息显然达不到目的，必须借助其他的数学方法从众多的信息中筛选出建模的特征光谱数据。应用Bi方法筛选出最能反映远安黄茶品质的子区间后建立PLS模型，经比较表2和表3的模型结果，应用Bi-PLS方法建立的远安黄茶品质NIRS模型，不仅建模的光谱数据少（占全部光谱数据的41.80%），而且模型的效果还得到了有效提升（RMSECV下降了15.83%）。筛选的特征光谱区间分别为9 003.2～7 497.9 cm^{-1}、6 101.7～5 449.8 cm^{-1}和4 601.3～4 246.5 cm^{-1}。在3个子区间范围内，均没有包含水的—OH近红外光谱（6 900 cm^{-1}和5 100 cm^{-1}）信息，避免了H_2O峰的高吸收对模型的影响。在已有研究中，与茶品质有关的内含成分主要为茶多酚、可溶性糖、咖啡碱、游离氨基酸和可溶性蛋白质等物质[24]。其中，茶多酚与茶苦涩味的形成有一定的关系，具有一定的收敛性；可溶性糖可增强茶汤的黏稠度，改善茶汤的滋味；咖啡碱与茶汤的苦味有关，游离氨基酸和可溶性蛋白质可提高茶汤的鲜味。上述这几类物质主要含有苯、酚羟基、—CH_x、—C＝O、—COOH和—NH_x等官能团，在近红外光谱区间，6 000 cm^{-1}附近是苯官能团的组合频信息区域，主要反映了茶多酚的光谱信息；5 555～5 882 cm^{-1}是—CH_x一级倍频信息区域，8 264～8 696 cm^{-1}是—CH_x二级倍频信息区域，4 545～4 500 cm^{-1}和6 666～7 690 cm^{-1}是—CH_x组合频信息区域，主要反映了单糖等葡萄糖和咖啡碱等茶叶碱以及茶氨酸等游离氨基酸的光谱信息；4 760～4 445 cm^{-1}是—C＝O组合频信息区域，

4 504 cm^{-1}附近吸收峰是 C—H 伸缩振动和C=O 伸缩振动的组合频信息区域，主要反映了咖啡碱以及单糖的光谱信息；4 630 cm^{-1}和 4 695 cm^{-1}附近是 C—H 伸缩振动和 C=O 伸缩振动的组合频信息区域，4 525 cm^{-1}附近是—NH_x一级倍频信息区域，4 613～4 587 cm^{-1}区域是 N—H 弯曲二级倍频与 C=O伸缩振动的组合频信息区域，主要反映了游离氨基酸、蛋白质和咖啡碱等茶叶碱的光谱信息[25]。但是在光谱区间中，与酚羟基有关的信息较少，可能是由于黄茶在闷黄过程中茶多酚的酚羟基极易被氧化为醌类物质或形成酮类物质的原因。可见，应用 Bi 算法筛选出的建模光谱区间可以较为准确地反映与远安黄茶品质密切相关的内含成分，具有较好的代表性，但是还有较多的光谱区间并没有与远安黄茶品质相关的官能团信息或者含有的信息非常微弱，将会影响模型的预测效果。因此，尝试应用 GA 算法来精准筛选与远安黄茶品质更加密切的数据点。

3.2　GA 精准提取特征光谱数据点

应用 GA 算法从 9 003.2～7 497.9 cm^{-1}、6 101.7～5 449.8 cm^{-1}和 4 601.3～4 246.5 cm^{-1} 3 个区间中精准提取了 75 个与远安黄茶品质密切相关的数据点（9 003.2～7 497.9 cm^{-1}中 40 个，6 101.7～5 449.8 cm^{-1}中 13 个，4 601.3～4 246.5 cm^{-1}中 22 个），其中最能反映远安黄茶品质内含成分官能团信息的数据点为 45 个（按照以上顺序分为 21、12、12 个，所占比例分别为 52.50%、92.31%、54.55%）。其中，9 003.2～7 497.9 cm^{-1}区间提取的特征光谱数据点主要反映的是—CH_x官能团的二级倍频和组合频信息，6 101.7～5 449.8 cm^{-1}区间提取的特征光谱数据点主要反映的是—CH_x官能团的一级倍频信息，4 601.3～4 246.5 cm^{-1}区间提取的特征光谱数据点主要反映的是—C=O 官能团组合频信息及—NH_x官能团一级和二级倍频信息，以上这些官能团与茶叶中的单糖、咖啡碱、茶氨酸和游离蛋白质有关，这几种成分都与远安黄茶品质密切相关[26]。其余 30 个光谱数据点主要反映的是木质素、淀粉、纤维素等多糖的 O—H 伸缩和 C—O 伸缩的组合频信息、酰胺组合频信息以及脂肪烃类物质的 C—H 伸缩和 C—C 伸缩的组合频信息。由于远安黄茶在干燥提香时，会发生蛋白质热裂解而生成游离胺[1]，而游离胺不稳定，会继续发生化学反应，生成糖胺类等。在评审远安黄茶品质时，用沸水冲泡黄茶并在浸泡过程中糖类物质以可溶解的单糖为主，但会有极少量的木质素、淀粉和纤维素等多糖物质溶解于茶汤中。上述物质虽然有助于提高茶汤的鲜味和甜味，但这些物质的含量极低，因此，它们的近红外光谱信息也就很微弱。这也是提取的 75 个特征光谱数据点中只有 45 个特征数据点可以较好地反映与黄茶品质有关的主要内含成分的官能团信息的原因。在今后研究中，进一步将茶叶中单糖、咖啡碱等理化成分与茶近红外光谱相结合，开展深层次茶品质评价研究。

本研究通过扫描获得了远安黄茶的近红外光谱，结合 Bi-PLS 算法和 GA-PLS 算法分别筛选了建模的特征光谱区间和特征光谱数据点，实现了远安黄茶品质的快速、无损、准确预测。模型具有较高的稳健性，可以准确地预测外部未知样品的品质得分（R^2=0.942，RMSEP=1.573）。但本研究在建立远安黄茶品质分数预测模型时，采用的样品均具有较好的品质，品质分数区间仅为 78.40～92.05。因此，为使模型具有较佳的稳健性和较广的应用范围，除了在建模时筛选能反映样品的特征光谱信息外，还需扩大样品的品质分数区间，补充一些品质分数稍低的样品用于建立预测模型，将得到很好的实际应用效果。

参考文献 References

[1] 宛晓春.茶叶生物化学[M].3 版.北京：中国农业出版社，2003. WAN X C.Tea biochemistry[M].3rd ed.Beijing：Chinese Agriculture Press，2003(in Chinese).

[2] 周继荣，陈玉琼，余志，等.鹿苑茶机械化加工技术研究[J].茶叶科学，2005，25(2)：141-145.ZHOU J R，CHEN Y Q，YU Z，et al.Studies on the mechanizing technology of Luyuan tea[J]. Journal of tea science，2005，25(2)：141-145(in Chinese with English abstract).

[3] KORTESNIEMI M，ROSENVALD S，LAAKSONEN O，et al. Sensory and chemical profiles of Finnish honeys of different botanical origins and consumer preferences[J]. Food chemistry，2018，246：351-359.

[4] HAN X，JIANG H，HAN L，et al.A novel quantified bitterness evaluation model for traditional Chinese herbs based on an animal ethology principle[J].Acta pharmaceutica sinica B，2018，8(2)：209-217.

[5] ZHU H K，YE Y，HE H F，et al.Evaluation of green tea sensory quality via process characteristics and image information[J]. Food and bioproducts processing，2017，102：116-122.

[6] WANG C,ZHANG C X,KONG Y W,et al.A comparative study of volatile components in Dianhong teas from fresh leaves of four tea cultivars by using chromatography-mass spectrometry,multivariate data analysis,and descriptive sensory analysis[J].Food research international,2017,100:267-275.

[7] ZHU J C,CHEN F,WANG L Y,et al.Evaluation of the synergism among volatile compounds in Oolong tea infusion by odour threshold with sensory analysis and E-nose[J].Food chemistry,2017,221:1484-1490.

[8] 中华人民共和国国家质量监督检验检疫总局,中国国家标准化管理委员会.中华人民共和国推荐性国家标准:茶 水浸出物测定 GB/T 8305—2013[S].北京:中国标准出版社,2014.General Administration of Quality Supervision,Inspection and Quarantine of the People's Republic of China,Standardization Administration of the People's Republic of China.National Standard (Recommended) of the People's Republic of China:tea:determination of water extracts content.GB/T 8305—2013[S].Beijing:Standards Press of China,2014(in Chinese).

[9] 陆婉珍.现代近红外光谱分析技术[M].北京:中国石化出版社,2000.LU W Z.Modern near infrared spectroscopy [M].Beijing:China Petrochemical Press,2000(in Chinese).

[10] GUILLEMAIN A,DÉGARDIN K,ROGGO Y.Performance of NIR handheld spectrometers for the detection of counterfeit tablets[J].Talanta,2017,165:632-640.

[11] 王胜鹏,滕靖,郑鹏程,等.青砖茶压制压力优化及GCG近红外快速检测模型建立[J].农业工程学报,2020,36(8):271-277.WANG S P,TENG J,ZHENG P C,et al.Optimizing processing pressure of Qingzhuan tea and development of GCG models for near infrared spectroscopy detection[J].Transactions of the CSAE,2020,36(8):271-277(in Chinese with English abstract).

[12] 周娇娇,吴潇扬,陈周,等.近红外光谱技术快速预测团头鲂新鲜度[J].华中农业大学学报,2019,38(4):120-126.ZHOU J J,WU X Y,CHEN Z,et al.Rapid prediction of *Megalobrama amblycephala* freshness by near infrared spectroscopy[J].Journal of Huazhong Agricultural University,2019,38(4):120-126(in Chinese with English abstract).

[13] WANG S P,ZHANG Z Z,NING J M,et al.Back propagation-artificial neural network model for prediction of the quality of tea shoots through selection of relevant near infrared spectral data via synergy interval partial least squares[J].Analytical letters,2013,46(1):184-195.

[14] 王胜鹏,郑鹏程,龚自明,等.青砖茶茶汤滋味品质的近红外快速无损评价[J].华中农业大学学报,2020,39(3):113-119.WANG S P,ZHENG P C,GONG Z M,et al.Rapid evaluation of taste quality for Qingzhuan tea soup based on near infrared spectroscopy[J].Journal of Huazhong Agricultural University,2020,39(3):113-119(in Chinese with English abstract).

[15] 吴瑞梅,赵杰文,陈全胜,等.近红外光谱技术结合特征变量筛选快速检测绿茶滋味品质[J].光谱学与光谱分析,2011,31(7):1782-1785.WU R M,ZHAO J W,CHEN Q S,et al.Determination of taste quality of green tea using FT-NIR spectroscopy and variable selection methods[J].Spectroscopy and spectral analysis,2011,31(7):1782-1785(in Chinese with English abstract).

[16] 刘小芳,刘钢,成红云.不同茶树品种黄茶的适制性研究[J].现代农业科技,2015(13):331,333.LIU X F,LIU G,CHENG H Y.Study on the suitability of yellow tea of different tea varieties [J]. Modern agricultural science and technology, 2015 (13):331,333(in Chinese).

[17] 周继荣,倪德江,陈玉琼,等.黄茶加工过程品质变化的研究[J].湖北农业科学,2004,43(1):93-95.ZHOU J R,NI D J,CHEN Y Q,et al.Study on quality variation during the yellow tea processing[J].Hubei agricultural sciences,2004,43(1):93-95(in Chinese with English abstract).

[18] 周继荣,陈玉琼,孙娅,等.鹿苑茶闷堆工艺研究[J].食品科学,2005,26(11):107-111.ZHOU J R,CHEN Y Q,SUN Y,et al.Studies on the piling technological effects on Luyuan yellow tea [J].Food science,2005,26(11):107-111(in Chinese with English abstract).

[19] HASHIMOTO T,GOTO M,SAKAKIBARA H,et al.Yellow tea is more potent than other types of tea in suppressing liver toxicity induced by carbon tetrachloride in rats[J].Phytotherapy research,2007,21(7):668-670.

[20] 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(3):413-419.

[21] SRINIVAS M,PATNAIK L M.Genetic algorithms:a survey [J].Computer,1994,27(6):17-26.

[22] 王春雷,卢彩云,陈婉芝,等.基于遗传算法和阈值滤噪的玉米根茬行图像分割[J].农业工程学报,2019,35(16):198-205.WANG C L,LU C Y,CHEN W Z,et al.Image segmentation of maize stubble row based on genetic algorithm and threshold filtering noise[J].Transactions of the CSAE,2019,35(16):198-205(in Chinese with English abstract).

[23] 严衍禄,陈斌,朱大洲,等.近红外光谱分析的原理、技术与应用[M].北京:中国轻工业出版社,2013.YAN Y L,CHEN B,ZHU D Z.Near infrared spectroscopy[M].Beijing:China Light Industry Press,2013(in Chinese).

[24] 叶飞,滕靖,龚自明,等.提香时间对远安黄茶理化品质的影响[J].现代食品科技,2017,33(11):96-101.YE F,TENG J,GONG Z M,et al.Effects of raising fragrant time on the physico-chemical quality of Yuan'an yellow tea[J].Modern food science and technology,2017,33(11):96-101(in Chinese with English abstract).

[25] 褚小立.化学计量学方法与分子光谱分析技术[M].北京:化学工业出版社,2011.CHU X L.Molecular spectroscopy analytical technology combined with chemometrics and its applications [M].Beijing:Chemical Industry Press,2011(in Chinese).

[26] 滑金杰，江用文，袁海波，等.闷黄过程中黄茶生化成分变化及其影响因子研究进展[J].茶叶科学，2015，35(3)：203-208.
HUA J J，JIANG Y W，YUAN H B，et al.Research progress on changes of biochemical components and influencing factors of yellow tea in the process of yellow tea suffocation [J].Tea science，2015，35(3)：203-208.

Fast and non-destructive quality evaluation of Yuan'an yellow tea based on near-infrared spectroscopy

WANG Shengpeng[1]，ZHENG Pengcheng[1]，GUI Anhui[1]，
TENG Jing[1]，LIU Panpan[1]，YE Fei[1]，GAO Shiwei[1]，MA Mengjun[2]，LIU Xiaoying[3]

1.*Institute of Fruit and Tea*，*Hubei Academy of Agricultural Sciences*，*Wuhan* 430064，*China*；
2.*Xianning Academy of Agricultural Sciences*，*Hubei Province*，*Xianning* 437100，*China*；
3.*Enshi Huazhishan Ecological Agriculture Co.*，*Ltd*，*Enshi* 445000，*China*

Abstract　Near infrared spectroscopy (NIRS) combined with a variety of algorithms was used to conduct rapid and non-destructive quality evaluation of Yuan'an yellow tea.The near-infrared spectra of 90 samples of Yuan'an yellow tea were obtained by scanning，and 11 different methods were used to preprocess the spectra to remove part of the noise information. Then the backward interval partial least squares (BiPLS) method was used to screen the characteristic spectral interval reflecting the quality of the sample and the genetic algorithm (GA) was used to accurately extract the characteristic spectral wavelength.Five NIRS prediction models of yellow tea quality were established.Finally，an attempt was made to analyze information of the spectral functional group.The results showed that the best spectral pretreatment method was multiple scattering correction. The characteristic spectral interval screened with BiPLS method was 9 003.2-7 497.9 cm^{-1}，6 101.7-5 449.8 cm^{-1} and 4 601.3-4 246.5 cm^{-1}，respectively.GA algorithm accurately screened 75 characteristic spectral data points.The Bi-GA-PLS model had the best robustness and predicted the quality of unknown samples ($R^2 = 0.951$，RMSEP = 1.57，RPD=5.27)，which was preliminarily realized to quickly and accurately forecast the quality of Yuan'an yellow tea.By analyzing the spectral information，45 spectral data points mainly reflected the information of functional group including —CH_x，C=O and —NH_x，representing the monosaccharide，caffeine，theanine and free protein closely related to the quality of the samples.The other 30 spectral data points mainly reflected the information of the functional group including O—H，amide bond and C—H and C—C stretching，representing lignin，starch and cellulose.

Keywords　Yuan'an yellow tea；quality evaluation；nondestructive evaluation；near infrared spectroscopy；backward interval partial least squares；genetic algorithm；artificial neural network

（责任编辑：陆文昌）