

适用于3类茶的定性分类及主要内含成分定量分析的近红外预测模型的建立

王梦东¹ 王胜鹏²

1. 武汉理工大学理学院, 武汉 430063; 2. 湖北省农业科学院果树茶叶研究所, 武汉 430071

摘要 以白茶、乌龙茶和红茶为研究对象, 力求同时实现对3类茶的定性鉴别和主要内含物含量的预测。采用因子化法建立定性预测模型, 该模型能够正确识别全部独立验证集样品; 采用偏最小二乘法结合主成分分析建立茶叶中含水量、茶多酚和咖啡碱的定量校正模型, 并用相关系数、交互验证均方差和预测均方差对模型进行评价。验证集中含水量、茶多酚和咖啡碱3个内含物成分定量预测模型的相关系数分别为0.9913、0.9057和0.9743。结果表明, 预测模型能实现同时对3类茶叶的定性分类和主要内含物含量预测的目的, 也达到了降低近红外预测模型成本的目的。

关键词 茶叶; 近红外光谱; 预测模型; 定性分类; 定量分析

中图分类号 TS 272.5 **文献标识码** A **文章编号** 1000-2421(2015)01-0123-05

近红外光谱(NIR)介于可见光区和中红外光区之间, 波长范围为0.75~2.50 μm , 波数范围为4 000~13 330 cm^{-1} 。其光谱信息来源于有机分子中的含氢基团(C-H、O-H、N-H)振动的合频与各级倍频的吸收, 是样品中有机分子含氢基团的特征振动信息, 化学信息量相当丰富^[1]。利用物质的近红外漫反射光谱实现对样品的无损、快速检测是近年来分析技术研究的热点。

自20世纪80年代以来, 近红外光谱技术与计算机程序分析相结合, 以其快速、简便和准确的特点成为第三代新兴分析技术而得到广泛应用, 目前已经被应用于食品、石化、烟草和医药等领域中。近红外分析技术应用于茶叶的定性和定量检测虽然已经有大量的研究报道^[2-7], 但尚未投入实际应用, 主要原因是茶叶产品种类繁多, 以一类茶叶为样品开发的预测模型往往不能应用于其他茶类, 建模代价大、设备开发成本高是困扰近红外检测技术应用于茶叶生产的关键问题。为此, 本文尝试建立能同时对白茶、乌龙茶和红茶进行定性分类和定量分析的近红外预测模型, 为近红外分析在茶叶生产中的实际应用提供借鉴。

1 材料与方法

1.1 样品收集

白茶茶样18个, 乌龙茶茶样22个以及红茶茶样32个, 分别产于浙江、福建和安徽等地。全部样本分成校正集和验证集, 其中校正集62个样品, 验证集10个样品(白茶2个, 乌龙茶3个, 红茶5个)。

1.2 样品内含成分的化学测定

根据国家标准规定的测定方法测定样品中3种内含物成分的含量。水分采用103 $^{\circ}\text{C}$ 恒重法测定^[8]; 茶多酚总量根据HPLC法测定^[9]; 咖啡碱是采用HPLC法测定^[10]。

1.3 样品的近红外光谱采集

采用美国Antaris型傅里叶变换近红外光谱仪(FT-NIR), 测量选用积分球漫反射光学平台; 光谱扫描范围12 500~4 000 cm^{-1} ; 分辨率4 cm^{-1} ; 检测器为InGaAs; 每个样品采集3次光谱, 每次扫描32次, 取3次采集光谱的平均值作为该样品的光谱。

1.4 样品近红外定性和定量预测模型的建立

采用TQ ANALYST软件建立样品定性和定量预测模型。建模前, 为了消除光谱的随机噪声和样品的不均匀性造成的影响, 先对原始光谱进行预

收稿日期: 2014-02-22

基金项目: 国家青年科学基金项目(31400586); 湖北省青年科学基金项目(2014CFB224); 湖北省农业科学院果茶所青年科学基金项目(GCJJ201301)

王梦东, 副教授。研究方向: 计算机与统计应用。E-mail: E8556@126.com

通信作者: 王胜鹏, 博士, 副研究员。研究方向: 茶叶加工与茶叶质量无损检测。E-mail: wwspp0426@163.com

处理,主要处理方法有:矢量归一化,多元散射校正,一阶导数+5、9、13、17、21、25 点平滑处理,二阶导数+5、9、13、17、21、25 点平滑处理,以及一阶、二阶导数(5、9、13、17、21、25 点平滑)+矢量归一化处理等;然后对波长进行优选,采用因子化法建立白茶、乌龙茶和红茶的近红外定性分析模型。同时,应用偏最小二乘法(PLS)和主成分分析法(PCA)建立 3 种茶叶样品的含水量、茶多酚含量和咖啡碱含量的近红外定量分析模型。建立模型时,将样品分为校正集和验证集,采用“留一法”的交互验证方法建立样品的校正集,用交互验证均方根方差(RMSECV)和相关系数分析所建模型的质量,并用独立验证集样品来检验模型的预测准确性。

2 结果与分析

2.1 茶叶样品中主要内含物的化学分析

采集样品中含水量、茶多酚含量和咖啡碱含量的化学分析结果见表 1。结果表明,化学分析方法测定结果标准偏差均小于 3%,说明方法的系统误差很小,分析结果稳定,适用于 3 类茶的定性分类及主要内含成分定量分析的近红外预测模型的建立。

表 1 3 种成分化学测定结果($n=72$)

Table 1 Contents of moisture, tea polyphenol and caffeine in white tea, oolong tea and black tea %

成分 Component	最小值 Min	最大值 Max	平均值 Mean	标准偏差 SD
水分 Moisture	2.75	10.28	6.89	1.89
茶多酚 Tea polyphenols	6.41	20.52	13.30	2.82
咖啡碱 Caffeine	1.86	4.75	3.10	0.75

2.2 样品的平均近红外光谱和建模波长的选择

在近红外光谱区茶叶样品的吸收重叠严重,吸收峰不明显。因此,建模的过程中对特征波长的选择尤为重要。图 1 为 3 种茶叶的平均近红外光谱图,在 $5\ 144\ \text{cm}^{-1}$ 和 $6\ 944\ \text{cm}^{-1}$ 波数附近有明显的吸收峰,这是茶叶样品中水分子 O-H 键伸缩振动的一级倍频和合频吸收区域,是水分吸收的敏感区,在建模的时候需尽量去除水分这 2 个特征波长区域,减少对模型的影响。因此,本文选择了 $7\ 400\sim 9\ 900\ \text{cm}^{-1}$ 范围内的光谱进行建模。

2.3 定性分类

1) 白茶、乌龙茶和红茶定性分类模型的建立。应用 TQ ANALYST 建立了白茶、乌龙茶和红茶的

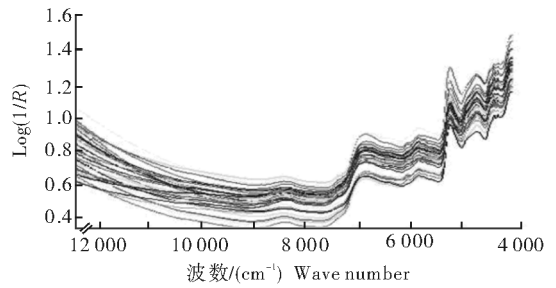


图 1 白茶、乌龙茶和红茶的平均傅里叶近红外光谱

Fig. 1 NIR reflectance spectra (log 1/R) from 12 500 to 4 000 cm^{-1} obtained by Bruker MPA

定性分析模型。经过多次优化后,对 $7\ 400\sim 9\ 900\ \text{cm}^{-1}$ 范围内的波长进行二阶导数(21 点平滑)+矢量归一化预处理,应用因子化法计算 3 类茶叶的阈值 T 和平均光谱之间的距离 D ,得出 S 值。当 $S>1$ 时,表明 3 类茶叶之间已经被完全分开。

表 2 白茶、乌龙茶、红茶定性分析模型结果

Table 2 The result of qualitative model of white tea, oolong tea and black tea

组 1 Group 1	组 2 Group 2	S	T1	T2	D
乌龙茶 Oolong tea	白茶 White tea	1.061 8	0.600 7	0.730 7	1.413 7
白茶 White tea	乌龙茶 Oolong tea	1.061 8	0.730 7	0.600 7	1.413 7
红茶 Black tea	白茶 White tea	1.077 0	0.582 3	0.730 7	1.414 1
红茶 Black tea	乌龙茶 Oolong tea	1.195 2	0.582 3	0.600 7	1.414 0

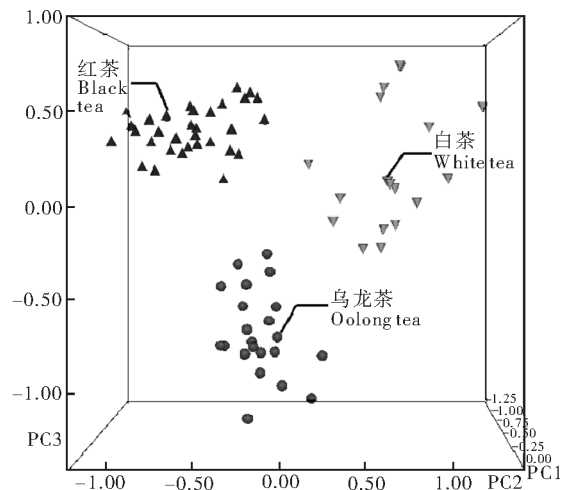


图 2 白茶、乌龙茶、红茶主成分空间分布图

Fig. 2 Three-dimensional picture of PC1, PC2 and PC3 of white tea, oolong tea and black tea

由表 2 可知,3 类茶叶之间的 S 值均大于 1,表明 3 类茶叶之间的平均距离(D)均大于它们的阈值之和,彼此之间没有出现交叉现象,是完全分开的。由图 2 可知,不同茶类之间界限明显,有明显的聚类现象,表明因子化法是识别茶类的一种有效方法。

2)定性分析模型的验证。利用独立验证集对建立的白茶、乌龙茶和红茶定性分析模型进行验证(表 3)。当样品与模型之间的匹配值小于某一类茶

表 3 独立验证集样品对定性分析模型的验证结果¹⁾

Table 3 The validation results for the qualitative model of white tea, oolong tea and black tea			
编号 No.	匹配值 Comparability	T	组 Group
W1	0.616 2	0.730 7	W ✓
	1.895 5	0.582 3	B
	2.016 9	0.600 7	O
W2	0.364 5	0.730 7	W ✓
	1.136 8	0.582 3	B
	1.493 9	0.600 7	O
O1	0.402 1	0.600 7	O ✓
	1.039 9	0.582 3	B
	1.389 2	0.730 7	W
O2	0.038 6	0.600 7	O ✓
	1.411 8	0.582 3	B
	1.445 9	0.730 7	W
O3	0.189 3	0.600 7	O ✓
	1.546 8	0.730 7	W
	1.599 7	0.582 3	B
B1	0.145 7	0.582 3	B ✓
	1.285 5	0.730 7	W
	1.412 8	0.600 7	O
B2	0.424 7	0.582 3	B ✓
	1.006 7	0.600 7	O
	1.161 3	0.730 7	W
B3	0.376 4	0.582 3	B ✓
	1.223 7	0.600 7	O
	1.598 0	0.730 7	W
B4	0.276 3	0.582 3	B ✓
	1.381 6	0.600 7	O
	1.629 0	0.730 7	W
B5	0.358 3	0.582 3	B ✓
	1.524 2	0.600 7	O
	1.757 0	0.730 7	W

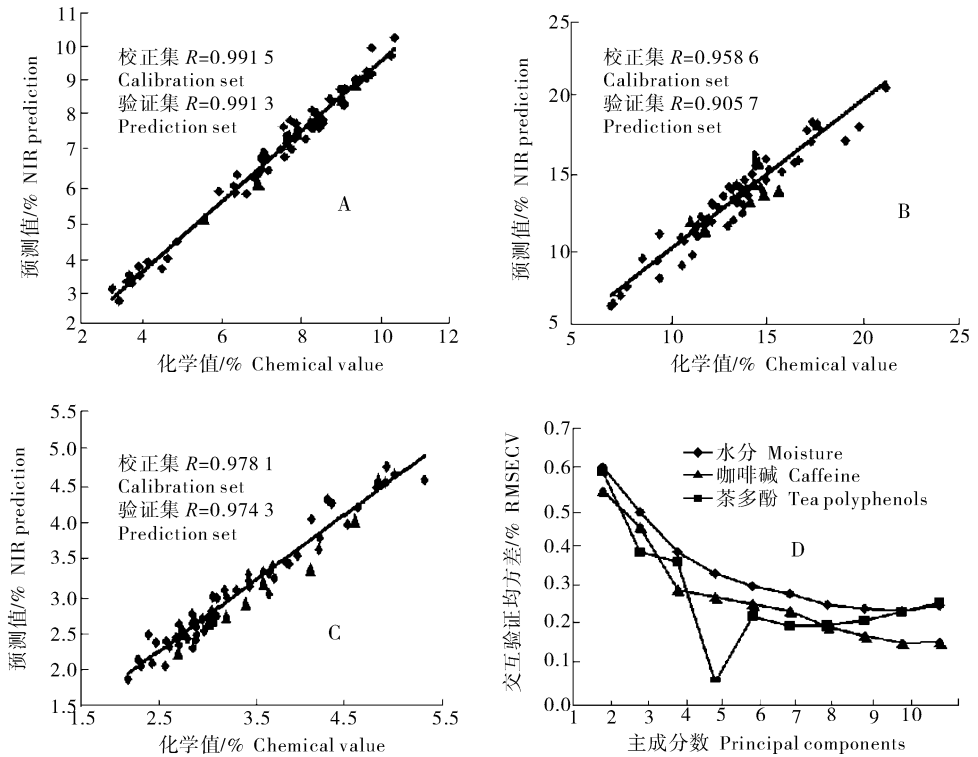
1) W:白茶样品 White tea; O:乌龙茶样品 Oolong tea; B:红茶样品 Black tea; 匹配值:样品与定性模型之间的距离 Match value: the distance between the sample and the qualitative model.

叶的阈值时,可以认为此样品就属于这一类茶叶。从表 3 可以看出,验证集中白茶、乌龙茶和红茶样品与定性分析模型的匹配值均小于它们各自茶类的阈值,表明验证集中的白茶、乌龙茶和红茶样品均被正确地判断归属于白茶类、乌龙茶类和红茶类。因此,基于傅里叶近红外光谱预测模型定性识别 3 种茶叶种类的方法是可行的。

2.4 定量分析

为了提高模型的预测准确性,在应用偏最小二乘法(PLS)和主成分分析法(PCA)建立 3 类茶叶的含水量、茶多酚和咖啡碱定量分析模型前,对光谱进行预处理。经过优化后,最终分别采用一阶导数(17 点平滑)+矢量归一化、二阶导数(17 点平滑)和多元散射校正的预处理方法。校正集定量模型采用交互验证的方法建立,并用交互验证均方差(RMSECV)和相关系数(R)和预测均方差(RMSEP)对模型进行评价,建立的含水量、茶多酚和咖啡碱近红外定量模型结果见图 3。

图 3 中直线表明的是近红外模型的预测值与化学测定值之间的关系,当图中样品点越靠近这条直线,说明近红外预测值越接近化学测定值,表明模型预测的准确度越高。图 3A 是含水量校正集模型,相关系数(R)为 0.991 5,交互验证均方差(RMSECV)为 0.234,用独立验证集样品来检验此模型的预测准确度时,得到相关系数(R)为 0.991 3,预测均方差(RMSEP)为 0.246;图 3B 是茶多酚总量校正集模型,相关系数(R)为 0.958 6,交互验证均方差(RMSECV)为 0.836,用独立验证集样品来检验模型的预测准确度时,得到验证集相关系数(R)为 0.905 7,预测均方差(RMSEP)为 0.875;图 3C 是咖啡碱校正集模型,相关系数(R)为 0.978 1,交互验证均方差(RMSECV)为 0.149,用独立验证集样品来检验模型的预测准确度时,得到验证集相关系数(R)为 0.974 3,预测均方差(RMSEP)为 0.198;图 3D 是 3 个校正集模型的最佳主成分数,最大 PC 值为 9,没有出现过拟合现象。表 4 是含水量、茶多酚、咖啡碱最佳校正集与预测集模型的建模结果。图 3 和表 4 说明了基于傅里叶近红外光谱法(FT-NIR)建立的含水量、茶多酚和咖啡碱含量的校正集模型达到了较高的预测准确度,可以用于白茶、乌龙茶和红茶样品相应内含物含量的检测。



A: 含水量 Humidity content; B: 茶多酚 Tea polyphenols; C: 咖啡碱 Caffeine; D: 主成分数、交互验证均方差 PC and RMSECV; ■: 校正集样品 Calibration samples; ▲: 验证集样品 Validation set samples.

图 3 近红外模型预测值与化学测定值的关系及模型的最佳主成分数 (PC)

Fig. 3 Relationships between chemical analysis and NIR-PLS prediction showed by R , RMSECV and PC

表 4 含水量、茶多酚和咖啡碱含量最佳校正集与预测集模型的建模结果

Table 4 Results from the optimal model of moisture, tea polyphenols, caffeine in the calibration set and the prediction set

成分模型 Model	最佳主成分数 PC	校正集 Calibration set		验证集 Prediction set	
		R	RMSECV/%	R	RMSECV/%
含水量 Moisture	9	0.991 5	0.234	0.991 3	0.246
茶多酚 Tea polyphenols	6	0.958 6	0.836	0.905 7	0.875
咖啡碱 Caffeine	9	0.978 1	0.149	0.974 3	0.198

3 讨论

茶叶产品分为 6 大类,按生产规模的大小顺序排列分别是:绿茶、黑茶、乌龙茶、红茶、白茶和黄茶。绿茶和黑茶特征比较明显,一般消费者容易识别,黄茶、红茶和白茶同属于发酵茶,品质相对接近,本文尝试了将这 3 种品质相近的茶类放在一起建立定性和定量分析模型,得到了较好的预测准确度。

茶叶产品分为 6 大类,按生产规模的大小顺序排列分别是:绿茶、黑茶、乌龙茶、红茶、白茶和黄茶。绿茶和黑茶特征比较明显,一般消费者容易识别,黄茶、红茶和白茶同属于发酵茶,品质相对接近,本文尝试了将这 3 种品质相近的茶类放在一起建立定性和定量分析模型,得到了较好的预测准确度。

应用近红外分析技术定量测定某一类茶叶中的内含物含量已经有大量的文献报道,而在实际应用中,由于检测方法的适用范围过窄,检测设备的开发生产成本又非常高,因此,近红外检测技术实际应用于茶叶品质分析的设备至今尚未出现。如果能开发

出能同时适用于多种茶类的近红外品质分析技术,无疑将使设备和方法的开发成本大幅降低,使近红外设备应用于茶叶品质分析成为可能。考虑到绿茶和黑茶相对于其他茶类来说,品质差异较大,而乌龙茶、红茶和白茶同属于发酵茶,品质相对接近,本文尝试了将这 3 种品质相近的茶类放在一起建立定性和定量分析模型,得到了较好的预测准确度。

参 考 文 献

- [1] 牛智有,林新. 茶叶定性和定量近红外光谱分析方法研究[J]. 光谱学与光谱分析, 2009, 29(9): 2417-2420.
- [2] YAN S. Review: evaluation of the composition and sensory properties of tea using near infrared spectroscopy and principal component

- analysis[J]. *Journal of Near Infrared Spectroscopy*, 2005, 13(6): 313-325.
- [3] 廖步岩, 张正竹, 夏涛, 等. 近红外光谱分析技术在茶叶质量属性甄别中的应用[J]. *安徽农业大学学报*, 2009, 36(2): 287-291.
- [4] 徐立恒, 吕进, 林敏, 等. 茶叶中3类主要组分的近红外光谱分析作为茶叶质量的快速评定方法[J]. *理化检验-化学分册*, 2006, 42(5): 334-336.
- [5] 张正竹, 廖步岩, 阎守和, 等. 近红外光谱(NIRS)技术在茶叶品质保真中的应用前景[J]. *食品工业科技*, 2009, 30(2): 349-352.
- [6] CHEN Q S, ZHAO J W, GUO Z M, et al. Simultaneous analysis of main catechins contents in green tea (*Camellia sinensis* (L.)) by Fourier transform near infrared reflectance (FT-NIR) spectroscopy[J]. *Food Chemistry*, 2009, 113: 1272-1277.
- [7] 周健, 成浩, 王丽鸳. 近红外技术在茶叶上的研究进展[J]. *茶叶科学*, 2008, 28(4): 294-300.
- [8] 中华全国供销合作总社. GB/T8304—2002, 茶水分测定[S]. 北京: 中国标准出版社出版, 2002.
- [9] 中华全国供销合作总社. GB/T8313—2008, 茶叶中茶多酚和儿茶素类含量的检测方法[S]. 北京: 中国标准出版社出版, 2008.
- [10] 中华全国供销合作总社. GB/T8312—2002, 茶、咖啡碱测定[S]. 北京: 中国标准出版社出版, 2002.

Establishment the prediction model for qualitative classification and quantitative analysis of major components for three kinds of tea by near infrared spectroscopy

WANG Meng-dong¹ WANG Sheng-peng²

1. *Faculty of Science, Wuhan University of Technology, Wuhan 430063, China;*

2. *Institute of Fruit & Tea, Hubei Academy of Agricultural Science, Wuhan 430071, China*

Abstract In order to identify and predict major components for white tea, oolong tea and black tea in the same time, the 3 different tea products were collected, and the qualitative and quantitative models were established based on NIRs (near infrared spectroscopy). The qualitative classification model, which was established by factor-method, could identify full of the independent validation set samples. At the same time, partial least squares (PLS) algorithm and principal component analysis (PCA) were conducted on the calibration of regression model. The performance of the final model was evaluated according to root mean square error of cross-validation (RMSECV), root mean square error of prediction (RMSEP) and correlation coefficient (*R*). The correlations coefficients (*R*) in the prediction set were achieved as follows: 0.991 3 for moisture model, 0.905 7 for tea polyphenols model, and 0.974 3 for caffeine model. It was demonstrated that it was feasible to establish a “saving” qualitative and quantitative prediction model for 3 different kinds of tea products with the premise of accuracy.

Key words tea; NIR; prediction model; qualitative classification; quantitative analysis

(责任编辑: 陆文昌)