

基于近红外的 PLS 量化模型鉴定西湖龙井真伪的研究

周 健¹, 成 浩^{1*}, 贺 巍², 王丽鸳¹, 吴 迪¹

1. 农业部茶叶化学工程重点开放实验室, 中国农业科学院茶叶研究所, 浙江, 杭州 310008
2. 南京农业大学, 江苏 南京 210095

摘 要 对西湖龙井进行原产地精确鉴定是规范市场, 打击假冒伪劣的迫切需求。文章利用近红外技术对西湖龙井的真伪鉴定进行了研究, 提出了一种模型以进行西湖龙井鉴别的新方法。结果表明, 利用西湖龙井茶和其他地区以龙井加工工艺制成的扁形茶全区域的近红外原始光谱, 分别对其进行赋值, 采用 PLS 法建立了西湖龙井的预测模型(主成分数为 15), 通过预测值和西湖龙井的临界值进行比对实现了对西湖龙井的真伪的准确鉴定。对 70 份定标样品和 24 份外部验证未知样品鉴定结果的准确率都达到了 100%, 证明利用定量分析的 PLS 法建立的模型能有效准确地进行西湖龙井的真伪鉴定。

关键词 近红外; 光谱; 鉴定; 偏最小二乘法; 西湖龙井

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2009)05-1251-04

引 言

西湖龙井至今已有一千多年的历史, 位列中国十大名茶之首。2001 年, 国家质检总局正式对龙井茶实行原产地域保护, 将龙井茶定义为: 以“龙井”地名命名, 划定西湖产区、钱塘产区、越州产区 3 个产区, 用在原产地域范围内经认定的茶园生产的茶鲜叶, 并在原产地域内按《龙井茶》标准生产加工的绿茶。同时规定只有用杭州西湖产区的茶鲜叶生产的龙井茶才能称为“西湖龙井茶”。2005 年 7 月 15 日开始实施的《地理标志产品保护规定》使西湖龙井地理标志的保护得到进一步完善^[1]。但是目前西湖龙井受假冒伪劣产品的影响非常严重。西湖龙井和其他地区以龙井茶加工工艺制成的其他扁形茶在形态、滋味、香气等方面都很难区分^[2], 因此需要一种能快速进行西湖龙井鉴定的方法。

赵杰文等采用近红外光谱结合主成分-马氏距离模式识别方法区分了龙井、碧螺春、毛峰和铁观音四种中国名茶, 在 6 500~5 300 cm^{-1} 光谱范围, 通过 MSC 预处理方法, 以 8 个主成分建模, 对校正集样本和预测集样本的鉴别率分别达到 98.75% 和 95%^[3]; 陈全胜用近红外光谱结合 SIMCA 模式识别了龙井、碧螺春、祁红和铁观音四种茶叶, 发现在主成分数分别为 4, 5, 2 和 3 时, 四类模型的对未知茶叶样本的识别率分别是 90%, 80%, 100% 和 100%, 拒绝率全是

100%^[4]; 又用支持向量机(support vector machine, SVM)法成功地利用近红外对绿茶、红茶、乌龙茶实现了快速分类, 训练集和验证集的判别准确率都达到了 90% 以上^[5, 6], 证明了利用近红外技术在茶叶的品种分类上的可行性。

由于加工工艺的不同, 茶的品质成分差异很大, 因此不同种类茶的近红外光谱有较大的差异性, 往往比较容易进行区分。而西湖龙井和其他地区以龙井茶加工工艺制成的其他扁形茶, 因其加工工艺往往相同, 其光谱的相似度很高, 利用传统的判别方法如主成分分析法(PCA)等往往很难将其分开。李晓丽等在用近红外技术对茶叶种类进行判别研究时发现利用 PCA 可以较好地地区分羊岩勾青、雪水云绿、庐山云雾等茶叶种类, 但是很难区分西湖龙井和浙江龙井^[7]。本实验采用偏最小二乘法(PLS)建立定量的模型来进行西湖龙井真伪的定性鉴别, 取得了好的结果。

1 材料与方 法

1.1 实验材料

本试验采用的样品为杭州地区的西湖龙井以及浙江省的建德、金华、丽水、新昌、温州等不同产地以龙井茶加工工艺制成的其他扁形茶(以下简称“其他扁形茶”)共 94 份。其中西湖龙井样品 25 份, 其他地区的扁形茶样品 69 份。西湖龙井的样品直接现场取自杭州市西湖龙井产区的茶农及厂

收稿日期: 2007-11-18, 修订日期: 2008-02-22

基金项目: 国家科技基础条件平台项目(2005DKA1002)和浙江省自然科学基金项目(Y305096)资助

作者简介: 周 健, 1979 年生, 中国农业科学院茶叶研究所助理研究员 e-mail: zjph263@126.com

* 通讯联系人 e-mail: chenghao@mail.tricaas.com

家,其他地区的扁形茶样品也是直接到产地的生产厂家直接获取(表 1)。

Table 1 Origin of tea samples in research

西湖龙井		其他以龙井茶制作工艺的扁形茶	
产地	样品数量/个	产地	样品数量/个
杭州龙井村	9	浙江建德	6
杭州外大桥	6	浙江金华	8
杭州梅家坞	2	浙江温州	5
杭州龙坞慈母桥	5	浙江丽水	33
杭州翁家山	3	浙江新昌	17

1.2 样品处理

在进行近红外光谱的采集前,所有的样品均经过粉碎处理。称取大约 20 g 左右的茶样,放入中药粉碎机(DFT-50, 20 000 r·min⁻¹, 浙江林大机械厂生产)粉碎约 30 s 左右,之后将磨碎后的粉末过 40 目筛,准确称取 10 g 作为近红外的分析样品。

1.3 近红外光谱的采集

样品的近红外光谱采集在 IFS 28/N(Bruker, 德国)近红外光谱仪上进行,随机软件为 Opus Quant 2。光谱为 64 次扫描的平均值,扫描区域为 10 000~3 500 cm⁻¹。数据点的间隔为 3.857 cm⁻¹,光谱数据点为 1 946 个。采集时室温控制在 25 °C 左右,湿度保持稳定。

1.4 西湖龙井真伪鉴定模型的建立

随机选取西湖龙井样品 20 个,其他扁形茶样品 50 个,其中西湖龙井赋值为 1,其他扁形茶赋值为-1,采用内部交叉验证,以 PLS 建立西湖龙井的鉴定模型。以 0 为中间值进行真伪判定,其中通过模型的计算值<0 的为其他扁形茶,>0 的为正宗西湖龙井,采用剩下的 5 个西湖龙井和 19 个其他扁形茶样品为未知样品,检验模型的准确度。

1.5 数据分析

采用软件 Unscrambler95(CAMO, OLSO, 挪威)进行数据分析,包括光谱的预处理,PLS 建立模型和未知样品的鉴定等。

2 结果与分析

2.1 PCA 分析

西湖龙井和其他扁形茶的近红外光谱存在非常高的相似性(图 1)。图 2 横坐标代表试验样品的第一主成分得分值,纵坐标为样品的第二主成分得分值。经过比较发现,虽然主成分 1 和主成分 2 的累积贡献率都达到了 94%,但聚合度并不好,其分布不存在规律性,难以对西湖龙井的真伪进行鉴定,因此需寻找其他方法。

2.2 PLS 定量模型对西湖龙井的真伪鉴定

采用对西湖龙井和其他扁形茶分别赋值的方法,建立了相应的 PLS 定量分析模型。其中西湖龙井赋值为 1,其他扁形茶赋值为-1,以茶样的近红外光谱为 X 变量,赋值为 Y 变量,采用内部交叉验证方法,建立 PLS 模型。以 0 为区分

的临界值,>0 的定义为正宗西湖龙井,<0 的定义为其他扁形茶。

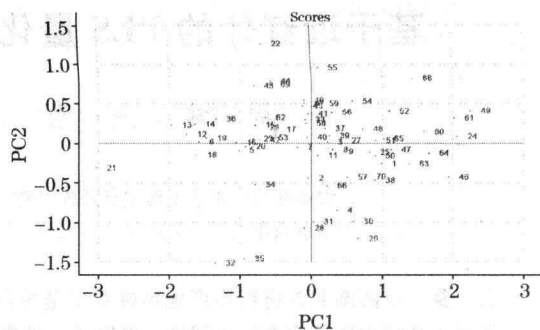


Fig. 1 Average spectrum of samples

1: Xihu longjing tea; 2: Other flat-shaped tea

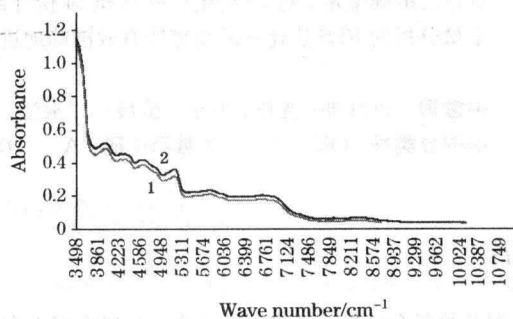


Fig. 2 PCA scores plots(PC1 ×PC2) for tea sample across the entire spectral region

1-20, Xihu longjing tea; 21-70, other flat-shaped tea

2.2.1 光谱区域的选择

针对茶样的近红外光谱的特点,将光谱分为下述区域: 3 500~5 300, 5 300~7 200, 7 200~11 000 和 3 500~7 200 cm⁻¹以及全光谱区域 35 000~11 000 cm⁻¹。在未经预处理的情况下比较建立的 PLS 模型对西湖龙井的真伪鉴定结果。以均方根误差 RMSECV 和相关系数 R² 为模型判别标准结果表明:采用 7 200~11 000 cm⁻¹,光谱区域建立的模型拟合度最低,而 RMSECV 值最高,而综合比较来看,采用全光谱区域,通过残余方差(Residual validation variance)分析,确定了最适主成分数为 15(图 3),建立的模型相关系数达到了 0.987,而 RMSECV 值仅为 0.102(详见表 2)。

Table 2 PLS models in different region of spectrum

	RMSECV	R ²
3 500~5 300 cm ⁻¹	0.218	0.942
5 300~7 200 cm ⁻¹	0.272	0.908
7 200~11 000 cm ⁻¹	0.560	0.615
3 500~7 200 cm ⁻¹	0.212	0.944
3 500~11 000 cm ⁻¹	0.102	0.987

2.2.2 模型验证

用建立的模型对 70 种定标样品进行鉴别,能很好地将西湖龙井和其他扁形茶区分开来,两者的模型预测值都分别

接近为其定义的值1和-1。按照以0为区分真假西湖龙井的临界值, >0的为西湖龙井, <0的为其他扁形茶的标准, 定标集样本的识别准确率达到了100%(见图4)。为了检测

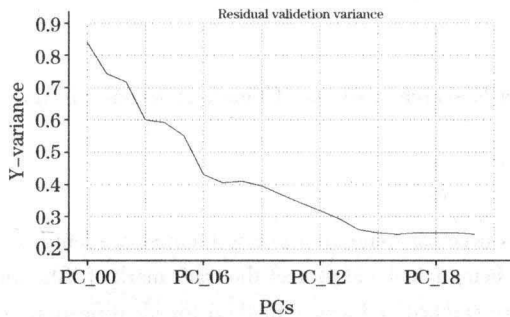


Fig. 3 Calculation of principal components (residual validation variance)

模型对未知样本预测的精确度, 将24个单独验证样品(包括5个西湖龙井样品和19和其他扁形茶样品)进行模型验证。结果表明, 模型对其是否为西湖龙井的识别准确率达到了100%(表3)。

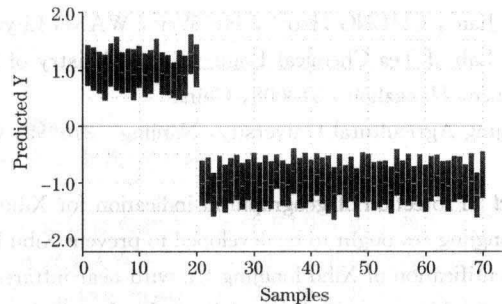


Fig. 4 Prediction of the model in sample set

Table 3 Prediction and validation of the model with unknown samples

种类	样品数量(个)	定义值	预测值	鉴定西湖龙井	判定样品数(个)	判定正确率
西湖龙井	5	1	0.182~0.965	是	5	100%
其他扁形茶	19	-1	-0.468~-2.166	否	19	100%

3 讨论

近几年来, 近红外技术在茶叶上的研究如定量分析(如茶多酚^[8]、咖啡碱^[9]、茶氨酸^[10]等)、定性鉴别等方面得到了很快的发展。本研究采用分别对西湖龙井和非原产地的以

龙井茶制作工艺制造的其他扁形茶分别赋值的方法, 以70个样本的近红外原始光谱为分析对象, 通过PLS结合内部交叉验证法建立了和西湖龙井的鉴别模型(模型主成分为15), 不仅对于定标集的70个样品实现了准确的鉴定(准确率100%), 而且对24个外部的未知样品的判别的准确率也达到了100%。本研究可作为其他名茶真伪鉴定的参考。

参 考 文 献

- [1] CAI En-ze(蔡恩泽). Foreign Business(外经贸), 2007, 3: 72.
- [2] SHEN Pei-he(沈培和). Newspaper of Tea(茶报), 2001, (3): 17.
- [3] ZHAO Jie-wen, CHEN Quan-sheng, ZHANG Hai-dong, et al(赵杰文, 陈全胜, 张海东, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2006, 26(9): 1601.
- [4] CHEN Quan-sheng, ZHAO Jie-wen, ZHANG Hai-dong, et al(陈全胜, 赵杰文, 张海东, 等). Journal of Food Science(食品科学), 2006, 27(4): 186.
- [5] Chen Quansheng, Zhao Jiewen, Fang C H, et al. Spectrochimica Acta Part A, 2007, 66: 568.
- [6] CHEN Quan-sheng, ZHAO Jie-wen, ZHANG Hai-dong, et al(陈全胜, 赵杰文, 张海东, 等). Acta Optica Sinica(光学学报), 2006, 26(6): 933.
- [7] LI Xiao-li, HE Yong, QIU Zheng-jun(李晓丽, 何勇, 裘正军). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2007, 27(2): 279.
- [8] Chen Quansheng, Zhao Jiewen, Huang Xingyi, et al. Microchemical Journal, 2006, 83: 42.
- [9] Luypaert J, Zhang M H, Massart D L. Analytica Chimica Acta, 2003, 478: 303.
- [10] SUN Yao-guo, LIN Min, LÜ Jin(孙耀国, 林敏, 吕进). Chinese Journal of Spectroscopy Laboratory(光谱实验室), 2004, 21(5): 940.

Identification of Xihu Longjing Tea by PLS Model Using Near-Infrared Spectroscopy

ZHOU Jian¹, CHENG Hao^{1*}, HE Wei², WANG Li-yuan¹, WU Di¹

1. Key Lab of Tea Chemical Engineering, Ministry of Agriculture, Tea Research Institute, Chinese Academy of Agriculture Science, Hangzhou 310008, China

2. Nanjing Agricultural University, Nanjing 210095, China

Abstract Protection of geographical indication for Xihu longjing tea was necessary and an accurate method of identification of Xihu longjing tea ought to be developed to prevent Xihu longjing tea from being faked and protect the trade mark of Xihu longjing tea. Identification of Xihu longjing tea with near-infrared spectrum was researched and a new method for identification of Xihu longjing tea by a quantitative model was developed. A new variance was first defined for Xihu longjing tea and other flat-shaped tea. And then the near-infrared spectrum without preprocess of tea was analyzed and partial least squares (PLS) was most suitably used to build the model (The number of PLS factors in this model is 9) for quantitative prediction. Compared with the critical numeral value, the result predicted by the PLS model was used to identify true Xihu longjing tea. The model was used to predict for 70 known samples which were used to build the quantitative model and 24 unknown samples. The recognition rate of 100% was achieved. So it is proved that the quantitative model by PLS was reliable and accurate to identify real Xihu longjing tea.

Keywords Near-infrared; Spectrum; Identification; PLS; Xihu longjing tea

(Received Nov. 18, 2007; accepted Feb. 22, 2008)

* Corresponding author