

# 基于近红外光谱技术的恩施玉露茶保存年份的快速无损鉴别

王胜鹏<sup>1</sup> 龚自明<sup>1</sup> 高士伟<sup>1</sup> 郑鹏程<sup>1</sup>  
滕靖<sup>1</sup> 郑琳<sup>1</sup> 卢素芳<sup>1</sup> 汪玲玲<sup>2</sup>

1.湖北省农业科学院果树茶叶研究所,武汉 430064; 2.湖北民族学院生物科学与技术学院,恩施 445000

**摘要** 为快速、准确、无损鉴别恩施玉露茶的保存年份,扫描在良好条件下连续保存5 a(2010—2014年)的100个恩施玉露茶,获得其近红外光谱,对光谱进行预处理,然后结合主成分分析法(PCA)和最小二乘支持向量机法(LS-SVM)建立绿茶保存年份的近红外光谱预测模型。结果表明,前3个主成分的累计贡献率为99.99%,验证集模型的决定系数( $R^2$ )为0.971 7,验证均方差(RMSEP)为0.255 0。初步实现了市售绿茶保存年份的快速鉴别,该方法也为其他茶类保存年份的判别提供参考。

**关键词** 近红外光谱;恩施玉露茶;保存年份;最小二乘支持向量机法

**中图分类号** TS 272.5 **文献标识码** A **文章编号** 1000-2421(2015)05-0111-04

茶叶作为世界“三大无酒精饮料”之首,备受大家的青睐。在我国,绿茶是最主要的茶类之一,其加工时间具有典型的季节性,特别是名优绿茶集中于清明节前生产加工,用于全年甚至多年消费,具有较长的储藏期。由于茶叶在储存期间容易受到水分、氧气、温度和光照等因素的不利影响,若保存不当,茶叶极易陈化变暗,香气变淡,经济价值明显下降。因此,为了更好的保存茶叶,使其经济价值得到有效的体现,通常将其保存在良好的条件下,避免其被氧化。绿茶通常保存条件为,含水量5%~6%、氧气含量低于0.1%、相对湿度低于30%、温度低于5℃,在此条件下保存绿茶具有良好的保鲜效果<sup>[1-2]</sup>,其经济价值也得到了最大程度的体现。但是,当这些保存良好的绿茶在茶店进行销售时,消费者很难通过肉眼鉴别其是否为新茶以及保存年份,因此,亟待寻求一种科学有效的快速鉴别绿茶保存年份的方法。

近红外光谱(near infrared spectroscopy, NIR)分析技术是一种快速、无损、绿色的分析方法,其与计算机程序分析相结合,具有快速、简便和准确的特点,已经被广泛应用于食品、石化、烟草和医药等领

域<sup>[3]</sup>。目前,国内外很多学者将光谱技术应用于茶叶的研究,一方面是对茶叶的内含成分如含水量、茶多酚、咖啡碱以及抗氧化能力进行快速测定以及茶叶等级的精确定级<sup>[4-10]</sup>等定量研究;另一方面主要用于茶叶种类的快速判别以及名优茶鉴定真伪<sup>[11]</sup>等定性应用研究。笔者以恩施玉露茶为研究对象,在良好保存条件下连续保存5 a(2010—2014年),使得茶叶仅凭肉眼无法判断其是否为新茶或陈茶,也无法直接判断出其保存的年份(年份值分别设定为1.000、2.000、3.000、4.000、5.000),样品粉碎后扫描其近红外光谱,然后对光谱进行主成分分析,最后应用最小二乘支持向量机法(LS-SVM)建立绿茶保存年份近红外光谱预测模型,对绿茶的保存年份进行快速、无损的判别,为消费者购买恩施玉露茶时提供一种极为便捷的判断方法,也为其他茶类的年份保存判别提供参考。

## 1 材料与方法

### 1.1 绿茶样品

样品来自湖北省恩施市某茶叶加工厂,茶叶加工时间为2010年4月,在含水量低于5%、氧气含

收稿日期:2015-04-22

基金项目:国家自然科学基金项目(31400586);国家现代茶产业技术体系建设专项(CARS-23);湖北省自然科学基金项目(2014CFB224);湖北省战略性新兴产业(支柱)产业人才培养(生物工程)项目

王胜鹏,博士。研究方向:茶叶加工/茶叶品质无损检测。E-mail: wwspp0426@163.com

通信作者:龚自明,研究员。研究方向:茶叶加工。E-mail: ziminggong@163.com

量低于 0.1%、相对湿度低于 30%、温度低于 5 °C 的避光条件下,将茶样在冷库中连续保存 5 a(2010—2014 年)。然后将样品从冷库中取出,将其粉碎后采用四分法,每个年份样品各 20 个,共 100 个样品。然后随机将样品分为校正集和验证集,其中校正集样品 75 个,验证集样品 25 个,用来检验模型的稳健性。

### 1.2 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;光谱扫描区间,4 000~10 000  $\text{cm}^{-1}$ ;分辨率,8  $\text{cm}^{-1}$ ;检测器,InGaAs;光谱数据分析软件为 TQ Analyst 9.4.45 和 Matlab R2007a(Mathworks, USA),模型结果用决定系数( $R^2$ )和验证均方差(RMSEP)表示。将全部 100 个样品每个采集 3 次光谱,每次扫描 64 次,取 3 次采集光谱的平均值作为该样品的最终光谱。光谱采集前,将仪器预热 1 h,保持室内温度和湿度基本一致,将样品倒入与仪器配套的旋转杯中,充分压实后采集光谱。

### 1.3 光谱预处理

由于茶叶光谱信息复杂,具有多种噪声信息,在建模之前必须对光谱进行预处理。经比较,采用移动平均平滑法,选用平滑窗口大小为 11,再进行多元散射校正(MSC)预处理,优选的波段为 4 500~9 000  $\text{cm}^{-1}$ 。

## 2 结果与分析

### 2.1 不同年份绿茶近红外光谱

从图 1 可以看出,保存年份不同的绿茶近红外光谱变化趋于一致,说明绿茶样品储藏条件较好,茶叶品质变化较小,没有出现明显的质变。但光谱在 6 500  $\text{cm}^{-1}$  和 5 100  $\text{cm}^{-1}$  附近吸收峰有较大的不同,说明保存年份不同绿茶样品间还是存在一定的

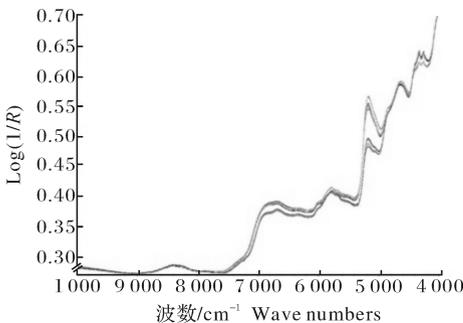


图 1 保存年份不同的绿茶样品近红外光谱

Fig.1 NIR spectra of the saved 1-5 years green tea

差异性,这为应用 LS-SVM 方法鉴别保存年份不同的绿茶奠定基础。

### 2.2 主成分分析

利用 MATLAB 软件对预处理后的 75 个校正集样品光谱数据进行主成分分析,求得第一主成分(PC1)贡献率为 94.77%,第一主成分(PC1)、第二主成分(PC2)和第三主成分(PC3)前 3 个主成分的累计贡献率为 99.99%,可以有效地代表校正集样品的光谱信息。

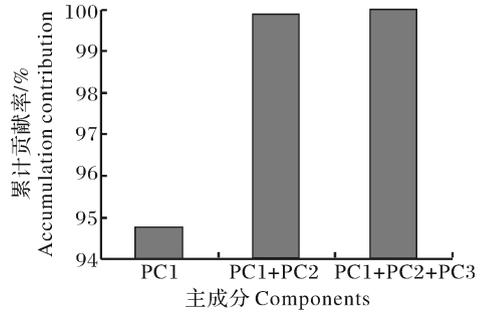


图 2 PC1、PC2 和 PC3 累计贡献率

Fig.2 Accumulating contribution of the top 3 components

### 2.3 LS-SVM 预测模型的建立

结合网格搜索法以及 10 等分交叉验证法优化处理超参数  $\gamma$  和  $\sigma^2$ ,以校正集交互验证均方误差(MSE)为目标函数<sup>[12-15]</sup>。寻优过程分为粗略优化和精细优化 2 个步骤,先通过粗略选择,再逐渐扩大搜索范围,以误差等高线来确立参数的初始范围,然后再逐步缩小搜索范围,精确确定参数值。其中,网格点“·”反映了第一步粗略搜索时格点搜索范围及步长,其中曲线反映了误差等高线,网络点“×”反映了第二步格点搜索范围及步长。根据上述原理,最终确定的最佳参数为  $\gamma=2\ 412.51, \sigma^2=0.278$ (图 3)。

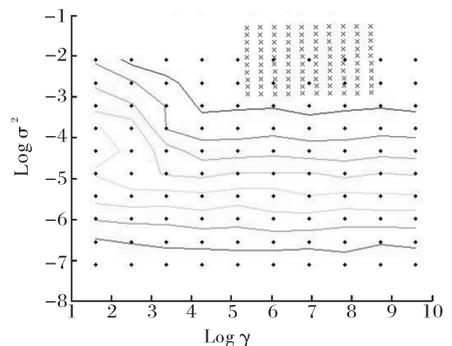


图 3 LS-SVM 中参数  $\sigma^2$  和  $\gamma$  的优化过程

Fig.3 Optimization of  $\sigma^2$  and  $\gamma$  for LS-SVM

图4是LS-SVM模型对验证集25个样品的预测结果( $R^2=0.9717$ ,  $RMSEP=0.255$ ),茶叶的年份预测值和年份真值之间相关性很高,表明建立的模型能够较为准确地预测绿茶的储存年份,达到了满意的结果,初步实现了绿茶保存年份的快速、无损鉴别。

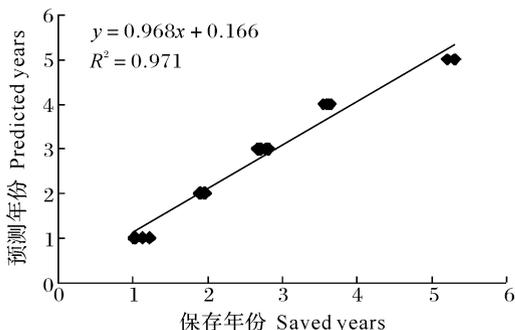


图4 LS-SVM方法预测结果

Fig.4 The best prediction model by LS-SVM method

### 3 讨论

在应用近红外光谱技术建立预测模型前,为筛选得到更多有用的光谱信息,提高信噪比,增强模型稳健性,需要去除高频随机噪声、基线漂移、样本不均匀和光散射等不利影响,对光谱进行预处理,包括平滑、多元散射校正(multiplicative scatter correction, MSC)、一阶导数(1st derivative)和二阶导数(2nd derivative)等,其中以移动平均平滑法与多元散射校正相结合效果最佳。最小二乘支持向量机方法具有非常强大的功能<sup>[16]</sup>,具有处理高维数据的强大能力,是多元建模分析中的一种非常有效的方法,它在非线性和空间识别上具有极强的优势,它克服了人工神经网络方法收敛难和不稳定等缺点,具有很高的可靠性。由于茶叶是茶树光合作用产物的主要储藏部位,成分非常复杂,含有大量与建模无关的光谱信息,光谱内部非线性明显。因此,为了减少输入变量,有效降低噪声信息,需先对光谱进行主成分分析,再以求得的主成分数为输入变量建立绿茶保存年份近红外光谱预测模型。而在应用RBF核函数建立LS-SVM预测模型时,最为关键的是要选择合适的参数: $\gamma$ 和 $\sigma^2$ 。其中, $\gamma$ 主要与建立模型时的复杂程度和逼近程度有密切的关联, $\sigma^2$ 主要与建立模型的精确度有密切的关联,本研究取得的最佳参数组合为 $\gamma=2412.51$ , $\sigma^2=0.278$ 。在以保存良好的不同储存年份(2010—2014年)的恩施玉露茶为

研究对象建立模型时,样品粉碎后扫描获得其近红外光谱,再对光谱进行预处理后进行主成分分析(前3个主成分累计贡献率为99.99%),最后应用最小二乘支持向量机法建立了恩施玉露茶不同储存年份的近红外光谱预测模型,该模型可以较为准确的预测验证集样品的储藏年份( $R^2=0.9717$ ,  $RMSEP=0.255$ ),初步实现了不同保存年份绿茶的快速、无损鉴别。

但是,本研究只针对了恩施市特定茶厂加工的恩施玉露茶,在建立预测模式的时候,建模样品并不是来自全部恩施玉露茶产区,因此,在实际应用时具有一定的限制。因此,当应用此模型前,尤其是在建立恩施玉露茶不同保存年份近红外光谱预测模型时,建议再增加一些建模样品,加入一些含有恩施玉露不同加工厂、鲜叶不同采摘海拔高度、不同加工季节、鲜叶不同采摘地点和不同土壤条件等信息的恩施玉露样品,扩充模型的信息,使得模型具有较广的使用范围,有利于增强建立模型的稳健性,更好的符合实际应用的需求,预测样品的保存年份越准确。同时,建立的模型也为其他一些茶类的保存年份鉴别提供了一种有益的参考,尤其是对普洱等黑茶类茶叶,更具有一定的参考价值。

### 参 考 文 献

- [1] 宛晓春.茶叶生物化学[M].3版.北京:中国农业出版社,2003.
- [2] 汤梦玲,海米梨,易雪峰,等.茶叶茶多酚及维生素C含量与储存时间相关性分析[J].大理学院学报,2014,13(2):58-60.
- [3] 陆婉珍.现代近红外光谱分析技术[M].2版.北京:中国石化出版社,2007:2.
- [4] YAN S H.Evaluation of the composition and sensory properties of tea using near infrared spectroscopy and principal component analysis[J].J Near Infrared Spec,2005,13(6):313-325.
- [5] CHEN Q S,ZHAO J W,ZHANG H D,et al.Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration[J].Analytica Chimica Acta,2006,572:77-84.
- [6] HE Y,LI X L,DENG X F.Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model[J].Food Engineering,2007,79:1238-1242.
- [7] YAN S H.NIR evaluation of the quality of tea and its market price[J].Spectroscopy Europe,2007,19(2):16-19.
- [8] 郭志明,赵杰文,陈全胜.特征谱区筛选在近红外光谱检测茶叶游离氨基酸含量中的应用[J].光学精密工程,2009,17(8):1839-1844.
- [9] 赵杰文,郭志明,陈全胜.近红外光谱法快速检测绿茶中儿茶素的含量[J].光学学报,2008,28(12):2302-2306.

- [10] 陈斌,叶静,颜辉.定标集样品数对茶叶近红外光谱分析精度的影响[J].江苏大学学报:自然科学版,2009,30(4):330-333.
- [11] 李晓丽,何勇,裴正军.一种基于可见-近红外光谱快速鉴别茶叶品种的新方法[J].光谱学与光谱分析,2007,27(2):279-282.
- [12] 高珏,王从庆.基于LS-SVM的苹果近红外光谱回归模型的研究[J].计算机测量与控制,2011,19(1):176-179.
- [13] 彭彦昆,黄慧,王伟,等.基于LS-SVM和高光谱技术的玉米叶片叶绿素含量检测[J].江苏大学学报:自然科学版,2011,32(2):125-129.
- [14] 徐惠荣,陈晓伟,应义斌.基于多元校正法的香梨糖度可见/近红外光谱检测[J].农业机械学报,2010,41(12):126-130.
- [15] 周竹,李小昱,李培武,等.基于GA-LSSVM和近红外傅里叶变换的霉变板栗识别[J].农业工程学报,2011,27(3):331-335.
- [16] 许禄.化学计量学[M].北京:科学出版社,1995.

## Identification of Enshi Yulu tea conserved years based on near infrared spectroscopy

WANG Sheng-peng<sup>1</sup> GONG Zi-ming<sup>1</sup> GAO Shi-wei<sup>1</sup> ZHENG Peng-cheng<sup>1</sup>  
TENG Jing<sup>1</sup> ZHENG Lin<sup>1</sup> LU Su-fang<sup>1</sup> WANG Ling-ling<sup>2</sup>

1. *Institute of Fruit & Tea, Hubei Academy of Agricultural Science, Wuhan 430064, China;*  
2. *College of Biological Science and Technology, Hubei University of Nationalities, Enshi 445000, China*

**Abstract** Enshi Yulu tea is a famous steaming green tea with a national geographical indication. Enshi Yulu tea has well-preserved appearance and color under strict storage conditions. To rapidly and non-destructively predict the quality of Enshi Yulu tea conserved years, 100 samples continuously stored 5 years (2010—2014) were studied with the near infrared spectroscopy. The least squares support vector machine combined with principal component analysis was used to establish the model for predicting the conservation years. The results showed that the cumulative contribution rate of the first three principal components, the coefficient of determination, and the root mean square error of prediction in the validation set was 99.99%, 0.971 7 and 0.255 0, respectively. The initial realization of the rapid identification of conservation years of commercially green tea is available. It will provide a useful reference for discriminating the conservation years of other teas.

**Key words** near infrared spectroscopy; Enshi Yulu tea; conservation years; least squares support vector machine

(责任编辑:陆文昌)