

# 基于近红外光谱的烤烟烟叶自动分组方法

章英<sup>1,2</sup>, 贺立源<sup>2\*</sup>

(1. 华中农业大学理学院计算机科学系, 武汉 430070; 2. 华中农业大学资源与环境学院, 武汉 430070)

**摘要:** 为了探索烤烟烟叶收购质量的无损检测技术, 提出了一种基于近红外光谱技术快速鉴别烟叶分组(部位、颜色)的方法。分析了近红外光谱技术应用于完整烤烟烟叶质量评价的可行性, 用不同波段范围、不同光谱预处理方法(多元散射校正 MSC、标准正态变量变换 SNV、微分光谱)和不同主成分因子数分别对烟叶部位和颜色分类结果的影响进行了对比分析, 分别建立了烟叶部位和颜色的定性判别模型。结果表明: 用判别分析(discrimant analysis, DA)方法在 1 101~2 395 nm 范围结合原始光谱建立的 DA 判别模型最优, 该方法对烟叶部位、烟叶颜色的校正集分类正确率均达 100%, 预测集分类正确率分别达到 98.57% 和 97.14%。说明所提出的方法具有很好的分组作用, 近红外光谱技术为烤烟烟叶收购质量等级评价提供了一种新方法。

**关键词:** 近红外光谱, 烤烟烟叶, 分组, 判别分析

doi: 10.3969/j.issn.1002-6819.2011.04.061

中图分类号: TN219

文献标志码: A

文章编号: 1002-6819(2011)-04-0350-05

章英, 贺立源. 基于近红外光谱的烤烟烟叶自动分组方法[J]. 农业工程学报, 2011, 27(4): 350-354.

Zhang Ying, He Liyuan. Auto-grouping method of flue-cured tobacco leaves based on near infrared spectra technology [J]. Transactions of the CSAE, 2011, 27(4): 350-354. (in Chinese with English abstract)

## 0 引言

目前烤烟烟叶收购过程主要依赖人工感官凭经验定级, 不但费时费力, 而且难以保证准确性和客观公正性。基于收购环节的烤烟烟叶自动分组研究大多采用常规图像处理方法, 借助烟叶的反射图像, 在可见光范围内研究以 RGB (red-green-blue) 或 HIS (hue-intensity-saturation) 颜色模型表述的烟叶光谱差异, 再用各种算法归纳等级特征。由于此类方法仅考察了烟叶的颜色、纹理、形状等外部品质, 难以将其内部品质信息与分组、分级相关联, 而且获取的光谱特征参数较少, 虽研究报道不少, 但进展不大<sup>[1-4]</sup>。

近红外光谱 (near-infrared spectroscopy) 属于现代结构分析技术, 具有不破坏样品化学结构、操作快速、简单和灵敏度高的特点, 已被广泛地应用于农产品品质分析<sup>[5-6]</sup>、食品品质检测<sup>[7-8]</sup>、烟草行业<sup>[9-10]</sup>等领域。李佛琳等<sup>[11]</sup>利用反射光谱建立了量化判别鲜烟叶成熟度模型并作了验证; 杜文等<sup>[12]</sup>研究了烟叶的近红外光谱数据通过软独立模式分类识别不同烟叶的方法; 刘剑君等<sup>[13]</sup>将支持向量机与红外光谱分析技术结合, 对烟叶分级进行了研究。

但是采用近红外光谱技术对烤烟烟叶“收购品质”进行自动分组、分级无损检测的, 目前尚无研究报道。

烟叶收购环节的质量控制按照“烤烟分级国家标准(42级)”(GB2635-1992)规定根据烟叶外观特征先分组、后分级, 分组后在同一组内再将烟叶划分 3~4 级。由于副组烟叶特征较明显且收购中占比较少, 故本研究着重于主组烟叶的划分。按部位和颜色主组烟叶划分为上部柠檬黄 BL、上部桔黄 BF、上部红棕 BR、中部柠檬黄 CL、中部桔黄 CF、下部柠檬黄 XL、下部桔黄 XF 等 7 组。本文应用近红外光谱分析技术开展对烤烟“整叶”的无损检测研究, 并探讨了烤烟烟叶部位、颜色自动分组快速检测方法, 对进一步开发烤烟烟叶自动化分级系统提供参考。

## 1 材料与方法

### 1.1 仪器设备

试验使用光谱检测设备是 Ocean Optics 公司的 NIR256-2.5 光纤光谱仪, 配套的 QBIF600-VIS-BX 白金级 Y 型分叉光导纤维探头, 该仪器光谱采样间隔 6 nm, 测定波长范围 900~2 500 nm。开机预热 1 h 后进行光谱扫描, 采样方式是漫反射, 采样软件是机器自带的 Spectra Suite。分析软件采用 ASD View Spec Pro, Unscramble V9 和 DPS (data procession system for practical statistics)。积分时间设置为 250 mm, 平滑度设置为 9, 平均次数为 3, 即对每个样点自动扫描 3 次取平均。

### 1.2 样品来源与数据采集

选取 2009 年 10 月云南省烟草烟叶公司提供的已由专家人工定级的烟叶。为保证试验结果具有代表性, 每组烟叶按 1~4 个等级随机选择 30 个样本, 7 组烟叶共

收稿日期: 2010-07-22 修订日期: 2011-03-04

基金项目: 国家科技支撑计划(2006BAD10A1304); 云南省烟草烟叶公司攻关项目(2009YN010)

作者简介: 章英(1978-), 女, 湖北枝江人, 讲师, 在职博士生, 研究方向为数字图像处理技术和农业信息工程。武汉 华中农业大学理学院计算机系, 430070。Email: zy@mail.hzau.edu.cn

\*通信作者: 贺立源(1951-), 男, 河南巩县人, 教授, 研究方向为资源环境信息工程。武汉 华中农业大学资源与环境学院, 430070。

Email: heliyuan@mail.hzau.edu.cn

210 个样本。按部位和颜色分类建模试验中，每个样本数据使用两次。每组 30 个样本随机分成预测集（20 个）和建模集（10 个），样本构成见表 1。

试验前首先用 2 块大玻璃对烟叶进行压制，以获得较好的平整性。扫描测定时，每片烟叶按图 1 所示，以主脉为轴对称划分出上下 2 个大区域 A 和 B，其中每个大区域再由叶尖至叶柄均分为 4 个小区域，即为 A1、A2、A3、A4 和 B1、B2、B3、B4 共计 8 个。

表 1 试验样本的组成  
Table 1 Compositions of test samples

		上部 (90 个)	中部 (60 个)	下部 (60 个)	颜色分类样本	
					建模集	预测集
柠檬黄(90 个)		30	30	30	60	30
桔黄(90 个)		30	30	30	60	30
红棕(30 个)		30			20	10
部位分 类样本	建模集	60	40	40		
	预测集	30	20	20		

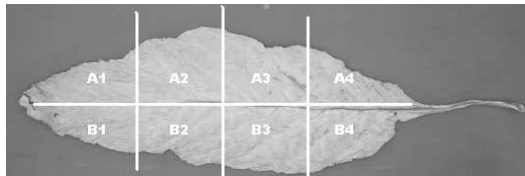


图 1 烤烟烟叶光谱测试区域划分图

Fig.1 Image of flue-cured tobacco leaves spectral testing areas

最后在暗室中进行光谱扫描（暗室温度为 18~22℃，相对湿度为 22%~25%，以卤素灯为唯一光源），每个小区域随机选取 5 个样点分别进行扫描（每点测 3 次后自动求平均），点与点间隔不超过 0.1~0.5 cm。光源与样品夹角 45°、光纤探测器与样品垂直呈 90°，探测器头部下端距离烟叶样品 4 cm。由光纤视场角 25°可计算出扫描覆盖范围的圆直径约为 1.77 cm。将每张烟叶共计 40 个样点的反射光谱求平均值得到每个样本的平均光谱（见图 2）。

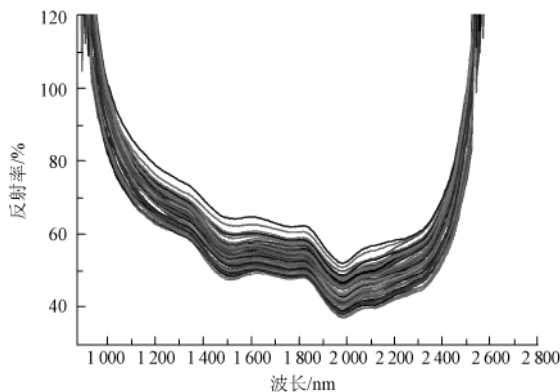


图 2 烟叶样本的近红外光谱图

Fig.2 Near-infrared spectra of tobacco leaf samples

### 1.3 光谱数据预处理

近红外光谱由于包含了噪声、环境闪素、仪器响应、人为操作误差等各种干扰因素，再加上烟叶样本表面不

均质特性，这些都对光的漫反射有一定影响。本文比较了 MSC（多元散射校正）、SNV（标准正态变量变换）、一阶导数和二阶导数等预处理方法对建模精度的影响。

由于不同光谱范围对校正模型判别精度影响很大，本文就以下 4 个波段范围对建模精度的影响进行了比较：长波近红外外段 1 101~2 395 nm、短波近红外 874~1 101 nm，TQ 软件推荐波段范围 1 101~2 588 nm，自选波段 1 101~2 395 nm。采用 TQ Analyst 软件中的 DA（discriminant analysis）法对烟叶部位、颜色分别进行判别分析研究。

## 2 结果与分析

### 2.1 近红外光谱用于完整烟叶检测分析

近红外漫反射光谱分析虽不需要对样品做化学处理，但是用于完整烟叶的检测时，由于烟叶自身叶面积跨度大的特点，光谱检测采样点的位置设定、采样点聚焦范围大小和烟叶表面平整度等因素都可能影响检测结果。由此探索了同一叶片不同测试区域的光谱特征及其差异、不同等级烟叶之间的光谱特征及其差异，明确了烟叶光谱采样点的代表性和应用于烤烟分级的可能性。

1) 对 210 个样本烟叶的近红外光谱数据进行分析，研究表明烤烟正组烟叶具有均质性，一个烟叶样本的光谱特征与测定区域关系不大，其峰谷的波长特征相同。单点测定反射率误差范围介于 2.38%~2.26%之间，区域反射率误差范围介于 2.36%~2.77%之间。该结果说明，采用光谱手段研究正组烟叶的光谱特征，可以任意选取烟叶的测定区域。2) 同一等级烟叶内不同烟叶样品间的光谱曲线中，峰谷特征波长位置值具有一致性，反射率差异极不显著。3) 不同等级烟叶之间的光谱特征差异比一片烟叶不同测试区域的光谱特征差异大得多，不同等级烟叶在同一波长处的反射率也表现为等级差异显著性高于区域差异。

### 2.2 不同波段对校正模型判别精度的影响

表 2 为不进行任何预处理，通过主成分分析选取 8 个主成分，在不同波段范围对 210 个样本分别建立部位、颜色校正模型的结果。

表 2 不同波段范围建立校正模型的正确分类率  
Table 2 Calibration result of different band regions using DA

波长/nm	样本数	部位判别		颜色判别	
		分类错误的样本数	正确分类率/%	分类错误的样本数	正确分类率/%
1 101~2 395	210	4	98.10	6	97.14
874~1 101	210	17	91.90	16	92.38
1 101~2 588	210	16	92.38	16	92.38
1 101~2 395	210	15	92.86	17	91.90

从表 2 中可以看出，与其他 3 个区域相比，长波近红外区域模型判别精度最好，对样本部位和颜色的正确分类率达 98.10%和 97.14%，这是因为烟叶样本主要吸收峰都在长波近红外区域。

2.3 不同预处理方法对校正模型判别精度的影响

表 3 对光谱范围为 1 101~2 395 nm, 使用 8 个主成分时, 对比分析了判别分析法结合一阶微分光谱及二阶微分光谱, 以及 MSC 和 SNV 平滑光谱所分别建立的部位、颜色判别模型分析结果。

表 3 不同的预处理方法建立判别分析校正模型结果  
Table 3 Calibration result of different preprocessed methods using DA

光谱数据类型	校正集(210 个样品)			
	部位识别		颜色识别	
	分类错误的样品数	样品正确分类率/%	分类错误的样品数	样品正确分类率/%
原始光谱	4	98.10	3	98.57
MSC+原始光谱	10	95.24	7	96.67
SNV+原始光谱	24	88.57	21	90.00
一阶微分光谱	80	61.90	74	64.76
二阶微分光谱	75	64.29	69	67.14
MSC+一阶导数	60	71.43	55	73.81
MSC+二阶导数	68	67.62	61	70.95
SNV+一阶导数	76	63.81	72	65.71
SNV+二阶导数	79	62.38	72	65.71

从表 3 可以看出, 没有经过任何数据预处理, 判别分析法建立校正模型就有较好的预测结果, 颜色和部位的样本正确分类率分别达到 98.10%和 98.57%。

2.4 不同主成分因子对校正模型判别精度的影响

主成分因子数也是影响DA法效果的一个重要因素。如果选取的主成分因子数过少, 将导致丢失较多有用的原始信息, 拟合不充分; 而主成分因子数太多, 则会加进许多噪声, 出现过度拟合现象, 不能得到稳定的预测模型。所建模型的预测误差会显著增大。从表4和表5可见, 当主因子数为12时, 即可对烟叶部位和颜色进行准确判别分析。此时, 提取12个主成分的贡献率分别已达全光谱贡献率的94.69%和94.72%, 而对分析区域光谱(1101~2395 nm)的贡献率则分别达99.85%和99.83%, 用210个样本用来建立校正模型时对烟叶部位、颜色分组的正确分类率为100%。之后, 随着主成分数的增加, 对分析区域的光谱累计贡献率增加非常缓慢, 故选用12个主成分进行DA判别分析。

表 4 主成分数在部位分类准确率和累计贡献率的影响  
Table 4 Influence of principal components on correct rate part's classification and cumulative contribution

主成分数	校正集(210 个样本)			
	对全光谱累计贡献率/%	对分析区域光谱累计贡献率/%	分类错误的样品数	样品正确分类率/%
8	90.61	99.76	3	98.57
10	93.50	99.82	0	100
12	94.69	99.85	0	100
14	95.32	99.86	0	100
16	97.11	99.87	0	100

表 5 主成分数在颜色分类准确率和累计贡献率的影响  
Table 5 Influence of principal components on correct rate color's classification and cumulative contribution

主成分数	校正集(210 个样本)			
	对全光谱累计贡献率/%	对分析区域光谱累计贡献率/%	分类错误的样品数	样品正确分类率/%
8	91.16	99.68	4	98.10
10	93.35	99.79	0	100
12	94.72	99.83	0	100
14	96.01	99.86	0	100
16	97.34	99.87	0	100

2.5 定性判别模型的建立和验证分析

2.5.1 校正模型的建立

在波段范围 1 101~2 395 nm, 不使用任何预处理方法, 用 12 个主成分分别建立烟叶部位和颜色的定性判别模型。图 3 所示为原始光谱的上部 B、中部 C、下部 X 叶 3 类之间距离的判别分析图, 图 4 所示为原始光谱的柠檬黄 F、桔黄 L、红棕 R 叶 3 类之间距离的判别分析图。3 个不同部位、3 种不同颜色的烟叶均可以得到较好的区分, 说明采用原始光谱结合判别分析法可以较理想的实现不同部位、不同颜色烟叶的分类。

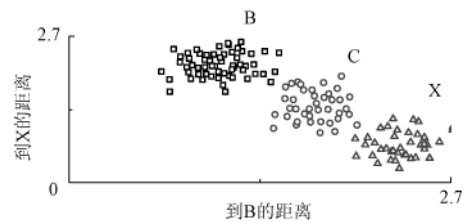


图 3 烟叶上部 B、中部 C 和下部 X 三类之间配对距离的判别分析图

Fig.3 Discriminant analysis plot of pairwise distance between three kinds of the tobacco leaf samples of B, C and X

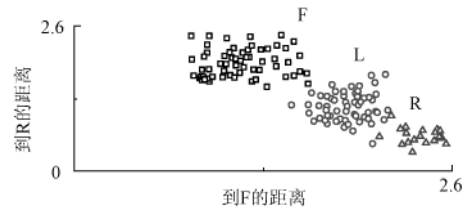


图 4 烟叶柠檬黄 F、桔黄 L 和红棕 R 三类之间配对距离的判别分析图

Fig.4 Discriminant analysis plot of pairwise distance between three kinds of the tobacco leaf samples of F, L and R

2.5.2 模型验证分析

采用判别分析法对烟叶部位预测集的 70 个样本的原始光谱进行预测分析, 有 1 个被误判, 正确识别率达到 98.57%。同样, 采用判别分析法对烟叶颜色预测集的 70 个样本的原始光谱进行预测分析, 有 2 个被误判, 正确识别率达到 97.14%。近红外原始光谱结合判别分析法可以较好地预测烟叶部位和颜色, 正确分组率达 97.86%。

### 3 结 论

1) 本文探讨了近红外光谱技术应用于烤烟烟叶收购环节的“整叶”品质无损检测评价, 试验表明烤烟正组烟叶具有均质性, 可采用单点或多点平均光谱反射率曲线代表整张烟叶的光谱特征。

2) 应用近红外漫反射光谱技术对烤烟烟叶的颜色和部位进行定性判别分析, 结果表明, 运用 DA 判别分析法, 光谱范围在长波近红外域 (1 101~2 395 nm), 主成分因子数为 12, 采用原始光谱建立的校正模型较优, 对校正集正确判别率为 100%, 预测集的正确判别率平均为 97.86%。

3) 应用近红外漫反射光谱的判别分析法对烤烟烟叶的部位和颜色进行判别是可行的, 仍需进一步深入研究, 如加大样本数量, 不同烟叶产地对烟叶类别预测精度的影响因素等, 以提高校正模型的精度和稳定性。

#### [参 考 文 献]

- [1] 张建平, 吴守一, 方如明. 农产品质量的计算机辅助检验与分级 (第 I 报) 烟叶外观品质特征的定量检验[J]. 农业工程学报, 1996, 12(3): 158—162.  
Zhang Jianping, Wu Shouyi, Fang Rumeng. Computer aided quality inspecting and grading for agricultural products (1st half) quantitative inspect ion of external quality characters of tobacco leaf[J]. Transactions of the CSAE, 1996, 12(3): 158—162. (in Chinese with English abstract)
- [2] 张帆, 方如明, 蔡健荣. 基于人工神经网络技术提取烟叶质量特征值[J]. 农业机械学报, 2000, 31(6): 61—64.  
Zhang Fan, Fang Rumeng, Cai Jianrong. Study of getting tobacco leaf weight based on neural network technology[J]. Transactions of the Chinese Society for Agricultural Machinery, 2000, 31(6): 61—64. (in Chinese with English abstract)
- [3] 周文, 韩力群, 李锐. 计算机图像处理技术在烤烟烟叶形状特征提取中的应用[J]. 烟草科技, 2000(1): 12—13,42.  
Zhou Wen, Han Liqun, Li rui. Application of computer image processing technology in drawing leaf shape character of flue-cured tobacco[J]. Tobacco technology, 2000(1): 12—13, 42. (in Chinese with English abstract)
- [4] 马文杰, 贺立源, 刘华波, 等. 成像环境因素对烟叶图像采集结果的影响及校正研究[J]. 中国农业科学, 2006, 39(12): 2615—2620.  
Ma Wenjie, He Liyuan, Liu Huabo, et al. The Influences of imagery environmental factors on the acquisition of flue-cured tobacco leaf images and correction[J]. Scientia Agricultura Sinica, 2006, 39(12): 2615—2620. (in Chinese with English abstract)
- [5] 朱大洲, 孟超英, 屠振华, 等. 不同类型近红外模型在苹果汁检测中的应用[J]. 农业工程学报, 2009, 25(增刊 2): 325—329.  
Zhu Dazhou, Meng Chaoying, Tu Zhenhua, et al. Application of different types of near infrared model to apple juice[J]. Transactions of the CSAE, 2009, 25(Supp2): 325—329. (in Chinese with English abstract)
- [6] 王多加, 周向阳, 金同铭, 等. 近红外光谱检测技术在农业和食品分析上的应用[J]. 光谱学与光谱分析, 2004, 24(4): 447—450.  
Wang Duoia, Zhou Xiangyang, Jin Tongming, et al. Application of near-infrared spectroscopy to agriculture and food analysis[J]. Spectroscopy and Spectral Analysis, 2004, 24(4): 447—450. (in Chinese with English abstract)
- [7] 汪庆平, 黎其万, 董宝生, 等. 近红外光谱法快速测定山核桃品质性状的研究[J]. 西南农业学报, 2009, 22(3): 873—875.  
Wang Qingping, Li Qiwan, Dong Baosheng, et al. Study on the prediction of oils of carya cathayensis sargent using near infrared reflectance spectroscopy[J]. Southwest China Journal of Agricultural Sciences, 2009, 22(3): 873—875. (in Chinese with English abstract)
- [8] 林涛, 于海燕, 应义斌. 可见/近红外光谱技术在液态食品检测中的应用研究进展[J]. 光谱学与光谱分析, 2008, 28(2): 285—290.  
Lin Tao, Yu Haiyan, Ying Yibin. Review of progress in application visible/near-infrared spectroscopy in liquid food detection[J]. Spectroscopy and Spectral Analysis, 2008, 28(2): 285—290. (in Chinese with English abstract)
- [9] 李向阳, 于建军, 刘国顺. 利用光谱反射率预测烤烟叶片烟碱含量[J]. 农业工程学报, 2008, 24(8): 169—173.  
Li Xiangyang, Yu Jianjun, Liu Guoshun. Predicting leaf nicotine content of flue-cured tobacco using hyperspectral reflectance[J]. Transactions of the CSAE, 2008, 24(8): 169—173. (in Chinese with English abstract)
- [10] 张峰, 刘江生, 刘泽春, 等. 近红外透射光谱技术用于烟用香精的品质控制[J]. 中国烟草学报, 2009(3): 12—16.  
Zhang Feng, Liu Jiangsheng, Liu Zechun, et al. Quality control in tobacco flavor by near-infrared transmittance spectroscopy[J]. China Journal of Tobacco, 2009(3): 12—16. (in Chinese with English abstract)
- [11] 李佛琳, 赵春江, 王纪华, 等. 一种基于反射光谱的烤烟鲜烟叶成熟度测定方法[J]. 西南大学学报: 自然科学版, 2008, 30(10): 51—55.  
Li Folin, Zhao Chunjiang, Wang Jihua, et al. A reflection spectrum-based method for determining the maturity grades of fresh flue-cured tobacco leaves[J]. Journal of Southwest University: Natural Science Edition, 2008, 30(10): 51—55. (in Chinese with English abstract)
- [12] 杜文, 易建华, 谭新良, 等. 基于近红外光谱的烟叶 SIMCA 模式识别[J]. 中国烟草学报, 2009, 15(5): 1—5.  
Du Wen, Yi Jianhua, Tan Xinliang, et al. SIMCA pattern recognition of tobacco leaves based on near infrared spectra[J]. China Journal of Tobacco, 2009, 15(5): 1—5. (in Chinese with English abstract)
- [13] 刘剑君, 申金媛, 彭丹青, 等. 基于 SVM 的烟叶光谱分级[J]. 通信技术, 2009, 42(11): 197—199.  
Liu Jianjun, Sheng Jinyuan, Peng Danqing, et al. Infrared Spectrum of tobacco grading based on support vector machine[J]. Communications Technology, 2009, 42(11): 197—199. (in Chinese with English abstract)

## Auto-grouping method of flue-cured tobacco leaves based on near infrared spectra technology

Zhang Ying<sup>1,2</sup>, He Liyuan<sup>2\*</sup>

(1. Department of Computer Science, College of Science, HuaZhong Agricultural University, Wuhan 430070, China;

2. College of Resources and Environment, HuaZhong Agricultural University, Wuhan 430070, China)

**Abstract:** In order to explore the nondestructive testing technology of flue-cured tobacco leaf on purchasing quality, the new method of fast classify tobacco leaf's group (parts, color) was proposed based on near infrared spectral technology. The feasibility of applying near infrared spectral technology to evaluate the complete flue-cured tobacco leaves quality was analyzed.. The prediction performances of different band range, different principal component numbers and different preprocessing methods of the spectra (multiplicative signal correction, standard normal variation, and derivative spectra) together with discriminant analysis (DA) were also investigated, and the calibration model was respectively established to classify the different parts and color of the flue-cured tobacco leaves. The research results for the tobacco leaf classification showed that DA calibration model using the parameters of band range between 1 101 and 2 395 nm combined with original spectra was optimal, which the correct percentage of classification on part and color of tobacco leaf was 100% for calibration sets, and it was 98.6% and 97.1% respectively for validation set. It is proved that the new method proposed in this study is capable to discriminate the parts and color of tobacco leaves with high accuracy. In addition, it might provide a new method to discriminate tobacco leaves group.

**Key words:** near-infrared spectral, flue-cured tobacco leaf, grouping, discriminant analysis