

DOI: 10.13652/j.issn.1003-5788.2020.11.021

基于近红外技术和偏最小二乘判别法对无花果成熟度的快速判别

Identification of fig maturity based on near-infrared spectroscopy and partial least square-discriminant analysis

周靖宇¹ 孙锐¹ 余多² 吕宇璇¹ 韩燕苓²ZHOU Jing-yu¹ SUN Rui¹ YU Duo² LV Yu-xuan¹ HAN Yan-ling²

(1. 齐鲁工业大学[山东省科学院]食品科学与工程学院, 山东 济南 250300;

2. 齐鲁工业大学[山东省科学院]数学与统计学院, 山东 济南 250300)

(1. Qilu University of Technology, Shandong Academy of Science, College of Food Science and Engineering, Jinan, Shandong 250300, China; 2. Qilu University of Technology, Shandong Academy of Science, College of Mathematics and Statistics, Jinan, Shandong 250300, China)

摘要:以无花果为试验对象,对其进行近红外光谱采集,并对其糖度、单果重、纵径、横径、硬度5个指标进行K均值聚类;根据光谱数据、主成分分析确定最优聚类效果的成分和各类别的指标分布构建偏最小二乘判别分析(PLS-DA)模型进行聚类判别,以实现对果实成熟度(幼果期、成长期、成熟期)分类的准确、快速、无损伤鉴别。结果表明,3种成熟阶段的无花果样品的糖度、单果重和硬度均具有显著性差异,成熟果和成长果与幼果的纵径和横径间具有显著性差异。根据PLS-DA判别模型累计训练集的分类正确率为99.59%,测试集的分类正确率为99.15%。说明主成分分析与光谱数据所建立的PLS-DA模型性能较好,对无花果成熟度的快速鉴别是有效且可行的。

关键词:近红外光谱;无花果;偏最小二乘判别分析;成熟度;鉴别

Abstract: In order to distinguish figs of different maturity accurately and quickly without damage, the samples of figs were collected by near-infrared spectroscopy, and K-means clustering was carried out for five indexes of fig sugar degree, single fruit weight, vertical diameter, horizontal diameter and hardness. According to the spectral data and the score chart of principal com-

ponent analysis, the components of the optimal clustering effect and the index distribution of each category were determined. Based on the PLS-DA model, the cluster discrimination model was constructed to achieve the purpose of fruit maturity classification. Through the analysis of the differences of the above five indexes, significant differences in soluble solids, single fruit weight and hardness of the three kinds of fig samples at the maturity stage were found, and significant differences between the vertical and horizontal diameters of mature and growing fruits and young fruits were also detected. According to PLS-DA discriminant model, the classification accuracy of training set was 99.59%, and that of test set was 99.15%. The results showed that the PLS-DA model based on principal component analysis and spectral data had good performance, and it could be used to identify the maturity of figs quickly.

Keywords: near infrared spectroscopy; fig; partial least square-discriminant analysis; maturity; identification

无花果(*Ficus carica* Linn.)属于桑科榕属,是世界上最早人工栽植的树种之一,是优良的经济果树^[1]。作为一种季节性食品,无花果是一些地区(如地中海地区)重要的饮食组成部分,可以鲜食或干制等^[2]。其营养丰富,富含纤维素、钾、铁和钙,果肉柔软,味甜,有健脾清肠、润肺利咽及提高免疫力等药用和抗氧化效果^[3-4]。成熟后的无花果易变软、腐败,不易贮藏,常温条件下难以长途运输。而无花果成熟的标志不明显,因此如何准确判别无花果成熟度对其销售、加工利用等极其重要^[5-6]。

基金项目:山东省自然科学基金(编号:ZR2017MC063);山东省重点研发计划(编号:2019GNC106139)

作者简介:周靖宇,男,齐鲁工业大学(山东省科学院)在读硕士研究生。

通信作者:孙锐(1978—),男,齐鲁工业大学(山东省科学院)副教授,博士。E-mail: sr@qlu.edu.cn

收稿日期:2020-05-03

目前,有关无花果的品质、活性物质以及加工工艺的研究较多,尤其是无花果栽培技术及多酚、黄酮等物质提取、抗氧化活性的研究^[7-9]。刘燕德等^[10]利用近红外漫反射光谱技术结合 Lab 颜色指标对西红柿成熟度进行检测,其预测集对 3 种成熟度的正判率分别为 67.86%, 50.00%, 100.00%。牛晓颖等^[11]采用 MPA 近红外光谱仪对李果实成熟度进行分类,其预测集的正判率为 96.30%。由于无花果为跃变型水果,成熟状态下采摘后会迅速软化,因此通常需要在果实成熟的早期阶段进行采摘。而无花果的采摘期通常是果农依据果实硬度、果皮颜色等经验确定^[12],同时果蔬种植属于劳动密集型产业,成本较高。Rungpichayapichet 等^[13]使用便携式 VIS-NIR 光谱仪,采用二阶导数预处理光谱对芒果成熟度进行分类,其准确率达 80% 以上;Pourdarbani 等^[14]根据果实光谱和颜色数据对富士苹果进行 4 个成熟阶段的预测,其精确度均 > 90%。以上研究表明,近红外光谱技术可作为果蔬品质无损评价的可靠方法,并能够建立有效稳定的预测模型。但当前关于无花果成熟度的预测分选、品质预测的研究较少,尚无更方便、有效和快速的检测方法。

理化指标如果实硬度、可溶性固形物等与感官指标如颜色、形状等是衡量无花果品质的重要指标,可在一定程度上体现无花果的成熟度。因此对于无花果成熟度的无损判别可分为基于光谱和图像技术的运用^[15-16]。图像技术只能根据果实表面特征进行分析,但果实在成长过程中外部通常是滞后于内部组织的变化^[17],且部分品种的无花果成熟与否颜色变化不大,可能会导致一些误判。而光谱技术能够反映部分果实内部信息,判别准确性较高,但目前关于无花果成熟度判别的研究报道较少。孙阳等^[18]采用图像识别技术,对摄取的果实图片通过预处理、分割和形态学处理达到区分不同成熟度无花果的目的。但在实际操作中,图像数据受光线、位置等条件影响,因此在此复杂背景下成熟果实的判别率会明显降低,而便携式近红外光谱仪采集光谱数据时则能有效减少这些因素的影响,降低模型的误判率。试验拟应用近红外光谱结合偏最小二乘判别分析(PLS-DA)法对无花果成熟阶段进行判别分析,以不同成熟阶段(幼果、成长果、成熟果)的无花果作为试验样本,利用近红外光谱仪进行光谱数据采集,采用随机划分确定校正集和预测集,根据校正集样品光谱和 3 种成熟阶段的分类变量建立 PLS-DA 判别模型^[19],然后对预测集进行判别,据此评价试验方法的效果,旨在为无花果成熟度分类的快速鉴别提供依据。

1 材料与方法

1.1 材料与仪器设备

1.1.1 材料

无花果:清晨于山东荣成市无花果种植基地(北纬

36°43'~37°27'、东经 122°09'~122°42')分别采摘幼果、成长果和成熟果 3 种阶段的无花果,果实均无损伤或腐烂,外观上大致相同,将无花果放于塑料泡沫包装盒中,迅速带回实验室于 4℃ 冰箱中贮藏,并于当日完成检测。

根据无花果果农的栽培经验和山东省林科院专家的建议,分为以下 3 种成熟阶段:幼果为果实生长时间 10~15 d,约为两成熟;成长果为生长时间 20~40 d,约为四至六成熟;成熟果为生长时间 50 d 以上,约为八至九成熟。

1.1.2 仪器设备

数字折光仪:PAL-1 型,日本 Atago 公司;

电子分析天平:FR124CN 型,奥豪斯仪器(上海)有限公司;

游标卡尺:CR2023 型,瑞士 Tesa 公司;

质构仪:TA.XT plus 型,英国 Stable Micro Systems 公司。

1.2 光谱采集

使用 MircoNIR 近红外光谱仪于 900~1 700 nm 内进行光谱数据采集。测量前,仪器需预热 1 h。光谱仪参数设置为:积分时间 10 ms,扫描次数 15 次,波长间隔 6.2 nm,共 125 个波长点。白天正常情况下测量,以 99% 反射白板的扫描光谱作为基准数据,每隔 30 个样品需要重新对反射白板进行扫描。沿无花果赤道随机选择 3 个不同位置获得光谱,取平均值作为该样品的原始光谱。若光谱明显区别于其他样品的近红外光谱,则在相同位置重新取样。

1.3 理化指标测定

1.3.1 糖度 每个样品从果实赤道切取相同大小的果肉,挤出果汁滴于折光仪镜面上,测量并记录读数,每个样品测 3 次,取平均值。

1.3.2 单果重 采用电子分析天平测定。

1.3.3 横径和纵径 采用游标卡尺测量。

1.3.4 硬度 采用质构仪,使用 P100 探头。下降速度与测试速度 1.0 mm/s,提升速度 1.0 mm/s,测试深度 30.0 mm,记录探入过程中所需应力,所得质构特征曲线的最大正峰值即为硬度。

1.4 聚类判别分析

采用 K-均值聚类对糖度、单果重、纵径、横径、硬度 5 个指标进行聚类,再使用 PLS-DA 对聚好的类别进行判别分析,建立无花果发育时期的判别模型。PLS-DA 将依据单果重、横纵径、糖度和硬度所聚类的不同类别对应的光谱数据作为输入变量 X,每个类采用数字作为标签,将标签数字转换为 5 位二进制数作为输出变量 Y,再将所输出变量 Y 转换为十进制数标签,判别分析使用 Python 的 Sk learn model selection 库中的 Train test split 方法随机

划分,所划分训练集为总数的67%,预测集为33%。

采用分类的精确率(P)和分类准确率(A)衡量判别模型性能,并分别按式(1)、(2)进行计算。

$$P = T_p / (T_p + F_p), \tag{1}$$

$$A = (T_p + T_n) / (T_p + T_n + F_n + F_p), \tag{2}$$

式中:

T_p ——将正类预测为正类数;

F_p ——将负类预测为正类数(误报);

T_n ——将负类预测为负类数;

F_n ——将正类预测为负类数(误报)。

2 结果与分析

2.1 无花果的近红外光谱

由图1可知,不同成熟期无花果近红外光谱曲线趋

势大致相似,均在1450 nm附近有明显的吸收峰。成熟果的主要吸收峰位于970.000,1119.235,1459.398 nm处,且1459.398 nm处较为强烈;成长果的主要吸收峰位于970,1160,1450 nm处,且1450 nm附近较为强烈;幼果的主要吸收峰位于990,1170,1459 nm附近,且1459 nm处较为强烈。970 nm处是由水和碳水化合物的吸收引起^[20],1119.24 nm处属于C—H键的第二泛音和O—H键的组合;1459.40 nm附近的吸收带对应于与水有关的O—H键拉伸的第二和第一振动泛音^[21]。

2.2 理化指标差异性分析

对3种成熟阶段的无花果共360个(幼果130个、成长果130个、成熟果100个)进行相关指标的测定。由表1可知,3种成熟阶段的无花果糖度、单果重、纵径和横

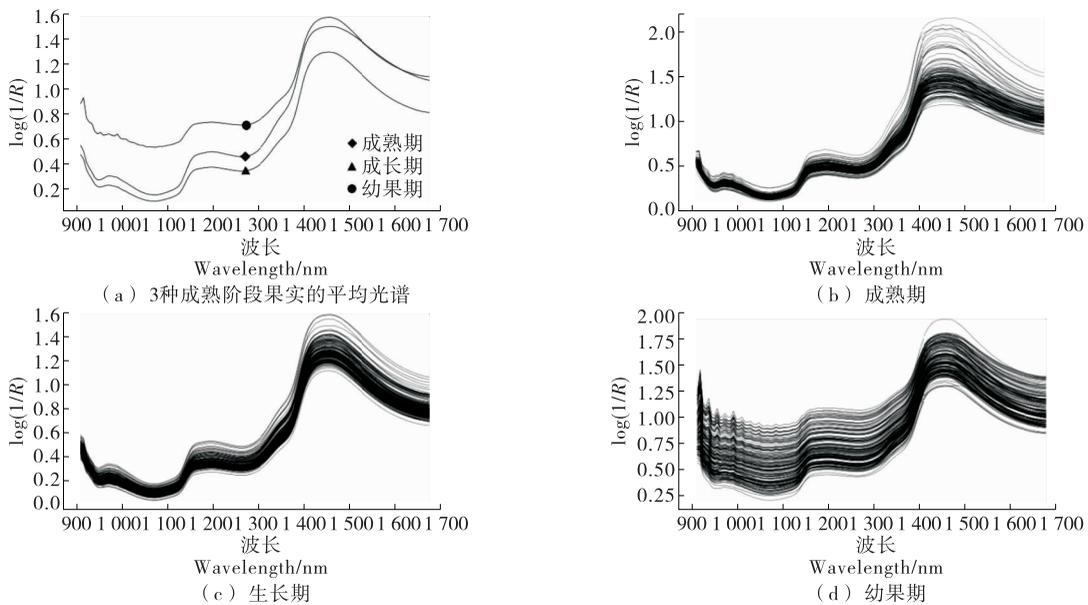


图1 无花果原始光谱图

Figure 1 The original spectrum of fig

表1 不同成熟阶段的无花果品质指标[†]

Table 1 Quality indexes of figs with different mature stage

成熟阶段	糖度/°Brix			单果重/g			纵径/cm		
	最小值	最大值	均值	最小值	最大值	均值	最小值	最大值	均值
成熟果	8.93	20.27	14.41±2.47 ^a	28.22	73.01	54.03±10.37 ^a	4.14	6.04	5.01±0.40 ^a
成长果	6.40	7.20	6.79±0.26 ^b	9.62	27.21	17.83±4.28 ^b	2.70	6.28	4.30±1.09 ^a
幼果	2.55	5.85	3.47±0.73 ^c	0.66	5.81	1.86±0.92 ^c	1.00	2.40	1.48±0.27 ^b

成熟阶段	横径/cm			硬度/g		
	最小值	最大值	均值	最小值	最大值	均值
成熟果	3.41	5.61	4.71±0.36 ^a	112.17	1990.02	591.4±364.15 ^c
成长果	3.10	6.10	4.59±0.83 ^a	3246.15	9231.47	5335.15±1000.03 ^a
幼果	1.98	3.50	2.73±0.34 ^b	967.43	4860.53	2514.79±609.46 ^b

[†] 字母不同表示差异显著(P<0.05)。

径 4 个理化指标的数值均随果实的成熟呈逐渐上升的趋势。幼果、成长果、成熟果糖度均值分别为 3.47, 6.79, 14.41 °Brix; 单果重分别为 1.86, 17.83, 54.03 g; 纵径均值分别为 1.48, 4.30, 5.01 cm, 横径均值分别为 2.73, 4.59, 4.71 cm。成长果的硬度值最高, 为 3 246.15 ~ 9 231.47 g, 幼果次之, 成熟果最低, 说明无花果成熟后变软、易损伤腐败和不易贮藏。3 种成熟阶段的无花果样品的糖度、单果重和硬度均具有显著性差异, 成熟果和成长果与幼果的纵径和横径间具有显著性差异, 说明糖度、单果重和硬度在 3 种成熟阶段时均有显著变化, 横径和纵径只在幼果与成熟果和成长果间有显著变化。

2.3 主成分分析

由表 2 可知, 5 个主成分能够提供样本所有的信息, PC1、PC2、PC3 3 个主成分累积反映了原始数据提供总量的 97.94%, 足以反映样本所包含的所有信息。因此, 将 PC1、PC2、PC3 3 个主成分作为评价样品质量的综合变量。

由图 2 可知, 不同成熟阶段的无花果样品按 PC1 有规律、清晰的分布, 说明主成分 1 对 3 种成熟阶段有很好的聚类作用; 而 PC2 对样品的得分分布影响不大, 幼果和成熟果的样品不能有效地区分开; 且成长果的样品得分较分散, 与上述测得的理化指标数据相似, 说明不同成熟阶段中成长时期的无花果理化指标含量变化幅度最大。

2.4 模型的建立与验证

由表 3 可知, 各类校正集和预测集的精确率均 > 0.99, 其中成熟果的校正集和预测集的无花果样品完全分类正确, 成长果和幼果的校正集完全正确。累计校

正集的分类正确率为 0.995 9, 预测集的分类正确率为 0.991 5, 正确率较高, 表明 PLS-DA 模型性能较好。孙阳等^[18]基于 Matlab 的无花果成熟度建立识别模型, 其正确率达 88.6%, 低于试验结果, 可能是其采用的针对于图像处理技术, 利用 RBG 颜色以及像素点分析判断果实是否成熟, 较近红外光谱所测得代表样品成熟与否的信息可能略少所致。

表 2 主成分的贡献率

Table 2 Contribution rate of principal components %

主成分	贡献率	累计贡献率
PC1	65.67	65.67
PC2	24.60	90.27
PC3	7.67	97.94
PC4	1.32	99.26
PC5	0.74	100.00

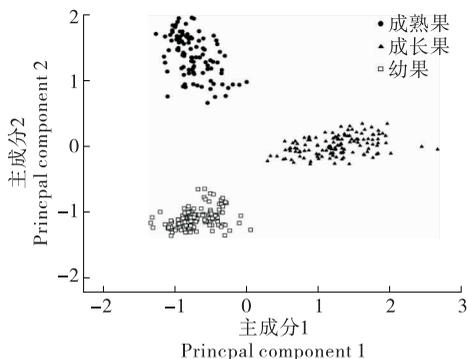


图 2 无花果样品的主成分得分图

Figure 2 PCA biplot scores plot for figs sample

表 3 PLS-DA 判别不同成熟度无花果样品

Table 3 The results of PLS-DA in identifying figs with different maturity

成熟度	校正集			预测集		
	总数	正确数	精确率/%	总数	正确数	精确率/%
成熟期	67	67	100.00	33	33	100.00
成长期	87	87	100.00	43	42	97.67
幼果期	87	86	99.85	43	43	100.00
总计	241	240	99.59	119	118	99.15

3 结论

对同一品种 3 种不同成熟阶段的无花果样本进行光谱采集, 根据糖度、单果重、纵径、横径和硬度 5 项品质指标的 K-均值聚类, 并结合以主成分为 3 的偏最小二乘判别分析构建判别模型, 根据预测集样本的预测准确度对模型效果进行验证。结果表明: 3 种成熟阶段的预测正确率分别为 100.00%, 97.67%, 100.00%, 预测集的预测准确率均达 95% 以上, 说明模型的预测效果较好。后续还需完善扩充样本种类, 对不同品种的无花果进行研究以

提高模型的适用性。

参考文献

[1] 孙锐, 贾明, 孙蕾. 世界无花果资源发展现状及应用研究[J]. 世界林业研究, 2015, 28(3): 31-36.
 [2] LADHARI A, GAALICHE B, ZARRELLI A, et al. Allelopathic potential and phenolic allelochemicals discrepancies in *Ficus carica* L. cultivars [J]. South African Journal of Botany, 2020, 130: 30-44.
 [3] 柴金珍, 黄远英, 袁根良, 等. 无花果的药理作用研究进

- 展[J]. 中成药, 2016, 38(8): 1 805-1 810.
- [4] 孙锐, 孙蕾, 贾明, 等. 山东引种无花果品种营养成分分析[J]. 经济林研究, 2014, 32(4): 63-67.
- [5] 廖亮, 李瑾瑜, 马红艳, 等. 贮藏温度和成熟度对新疆早黄无花果采后生理的影响[J]. 核农学报, 2016, 30(2): 282-287.
- [6] 孙锐, 孙蕾, 马金辉, 等. 各品种不同成熟度无花果质构特性分析[J]. 食品与机械, 2017, 33(2): 22-25, 30.
- [7] ABDEL-ATY A M, HAMED M B, SALAMA W H, et al. Ficus carica, Ficus sycomorus and Euphorbia tirucalli latex extracts: Phytochemical screening, antioxidant and cytotoxic properties[J]. Biocatalysis and Agricultural Biotechnology, 2019, 20: 101199.
- [8] PALMEIRA L, PEREIRA C, DIAS M I, et al. Nutritional, chemical and bioactive profiles of different parts of a Portuguese common fig (*Ficus carica* L.) variety[J]. Food Research International, 2019, 126: 108572.
- [9] 潘悠优, 花佩, 王允祥, 等. 无花果多糖提取、分离纯化及生物活性的研究进展[J]. 食品科学, 2016, 37(17): 289-295.
- [10] 刘燕德, 徐海, 孙旭东, 等. 西红柿成熟度的近红外漫透射光谱无损检测[J]. 激光技术, 2019, 43(1): 25-29.
- [11] 牛晓颖, 贡东军, 王艳伟, 等. 基于近红外光谱和化学计量学的李果实成熟度鉴别方法研究[J]. 现代食品科技, 2014, 30(12): 230-234, 125.
- [12] 孙锐, 孙蕾, 马金辉, 等. 不同成熟度无花果品质指标的变化分析[J]. 经济林研究, 2017, 35(2): 32-37.
- [13] RUNGPICHAYAPICHET P, MAHAYOTHEE B, NAGLE M, et al. Robust NIRS models for non-destructive prediction of postharvest fruit ripeness and quality in mango[J]. Postharvest Biology and Technology, 2016, 111: 31-40.
- [14] POURDARBANI R, SABZI S, KALANTARI D, et al. Automatic non-destructive video estimation of maturation levels in Fuji apple (*Malus Malus pumila*) fruit in orchard based on colour (Vis) and spectral (NIR) data[J]. Biosystems Engineering, 2020, 195: 136-151.
- [15] 李丽娜, 赵武奇, 曾祥源, 等. 苹果的质构与感官评定相关性研究[J]. 食品与机械, 2017, 33(6): 37-41, 45.
- [16] 毕智健, 张若宇, 齐妍杰, 等. 基于机器视觉的番茄成熟度颜色判别[J]. 食品与机械, 2016, 32(12): 133-136.
- [17] 黄玉萍, 刘英, 杨雨图, 等. 空间分辨光谱和可见/近红外光谱的番茄颜色等级判别[J]. 光谱学与光谱分析, 2019, 39(11): 3 585-3 591.
- [18] 孙阳, 白皓然, 初留珠, 等. 基于 Matlab 的无花果成熟度识别技术[J]. 计算机测量与控制, 2018, 26(10): 299-303.
- [19] 章林忠, 蔡雪珍, 方从兵. 近红外光谱定量和定性分析技术在鲜食葡萄果实无损检测中的应用[J]. 浙江农业学报, 2018, 30(2): 330-338.
- [20] CICCORITTI R, PALIOTTA M, AMORIELLO T, et al. FT-NIR spectroscopy and multivariate classification strategies for the postharvest quality of green-fleshed kiwifruit varieties[J]. Scientia Horticulturae, 2019, 257: 108622.
- [21] WEI Xuan, HE Jin-cheng, ZHENG Shu-he, et al. Modeling for SSC and firmness detection of persimmon based on NIR hyperspectral imaging by sample partitioning and variables selection[J]. Infrared Physics & Technology, 2020, 105: 103099.

(上接第 106 页)

5 结束语

烟梗种类繁多,形状各异。因产地、土壤、气候、年份等差异,其物理特性和耐加工性均有很大不同,对耐加工性显著不同的原料烟梗需精细化处理,对浸梗设备要求更高。试验设计只是其中一种改进方案和相关机构的创新设计,整体构造结构变化不大,却能有效改变烟梗处理的时间跨度,对浸梗设备设计开发具有很好的参考价值。试验改进方案仅通过过渡水槽进料端的两个引流口水槽对浸梗时间进行调节,虽然有效提高了浸梗机的回潮时间跨度,但是距离生产过程自动化、精细化、智能化等要求还相差甚远。如何让浸梗机浸泡装置自动识别耐加工性显著不同的原料烟梗,从而在生产过程中自动调节浸梗时间,以实现生产过程的自动化、精细化、智能化,还有待进一步研究。

参考文献

- [1] 刘钟祥, 罗登山, 雷樟泉, 等. 卷烟第 4 部分 感官技术要求: GB/T 5606.4—2005[S]. 北京: 中国标准出版社, 2005: 1-4.
- [2] 刘德强, 贾洋, 王乐军, 等. 烟丝结构对烟支卷制质量的影响[J]. 安徽农业科学, 2010, 38(32): 18 589-18 590.
- [3] 刘德强, 张风光, 王乐军, 等. 烟梗成丝的研究及应用[J]. 安徽农业科学, 2010, 38(33): 19 052-19 054.
- [4] 康金岭, 黄贤飞, 刘远涛, 等. 预处理工艺对烟梗回透率的影响[J]. 食品与机械, 2012, 28(6): 205-240.
- [5] 曾静, 李斌, 冯志彬, 等. 卷烟机剔除梗丝中含丝量的检测[J]. 烟草科技, 2012(8): 5-10.
- [6] 熊安言, 孙觅, 王镇增, 等. 浸梗工艺探讨[J]. 烟草科技, 2006(3): 9-14.
- [7] 杨继福, 黄兰, 卢幼祥, 等. 高压蒸梗对中性香味成分和感官质量的影响[J]. 广州化工, 2011, 39(17): 53-57.
- [8] 张志刚, 邵海民, 王振增, 等. 浸梗工艺的改进试验[J]. 烟草科技, 2007(2): 11-13.
- [9] 陈伟, 王镇增, 孙觅, 等. 正交试验法在浸梗机参数优化中的应用[J]. 烟草科技, 2008(7): 13-15.