

基于 K-means 聚类分析的不同种植区烤烟外观质量区域分类

杨明峰¹, 詹良², 魏春阳³, 李锋³, 王信民³

(1. 山东中烟工业有限责任公司技术中心, 青岛 266101; 2. 中国烟叶公司, 北京 100055;

3. 中国烟草总公司郑州烟草研究院, 郑州 450001)

摘要: 根据全国烤烟主产区生态特点, 采集了全国 112 个典型生态点初烤 C3F 烟叶样品, 运用快速聚类法(K-Means Cluster)分析了各烟草种植区烤烟外观质量区域特征及分布。结果表明, 5 大烟草种植区中部烤烟外观质量均可分为 3 类; “油分”和“色度”指标在北方烟草种植区烤烟样品间差异极显著, “叶片结构”指标在黄淮烟草种植区烤烟样品间差异极显著, “身份”和“色度”指标在长江中上游烟草种植区烤烟样品间差异极显著, “身份”和“油分”指标在东南烟草种植区烤烟样品间差异极显著, 烤烟外观各项指标在西南烟草种植区烤烟样品间差异均达极显著水平。

关键词: 烤烟; 外观质量; 快速聚类法; 区域分类

中图分类号: TS41⁺2

文章编号: 1007-5119(2012)02-0012-05

DOI: 10.3969/j.issn.1007-5119.2012.02.03

Regional Classification of Appearance Quality of Chinese Flue-cured Tobacco Based on K-Means Cluster

YANG Mingfeng¹, ZHAN Liang², WEI Chunyang³, LI Feng³, WANG Xinmin³

(1. Technology Center, China Tobacco Shandong Industrial Co., Ltd, Qingdao 266101, China; 2. China National Leaf Tobacco Corporation, Beijing 100055, China; 3. Zhengzhou Tobacco Research Institute of CNTC, Zhengzhou 450001, China)

Abstract: According to ecological characteristics of main flue-cured tobacco production areas, 112 of C3F tobacco leaf samples were collected from typical ecology areas. Regional characteristics and distribution of appearance quality of flue-cured tobacco from different tobacco planting areas were analyzed by K-means Cluster. The results showed that appearance quality of flue-cured tobacco middle leaves from five tobacco planting areas could be classified as 3 clusters respectively. Oil and color intensity difference in the leaves of north tobacco planting area were extremely significant. Leaf structure difference in the leaves of Huanghuai tobacco planting area was extremely significant. Body and color intensity difference in the leaves of the middle and upper reaches of the Yangtze river tobacco planting area were extremely significant. Body and oil difference in the leaves of southeast tobacco planting area were extremely significant. All the appearance quality indexes showed extremely significant difference in the samples of southwest tobacco planting Area. The results provide reference to cigarette industry.

Keywords: flue-cured tobacco; appearance quality; K-means cluster; regional classification

由于我国各烟草主产区气候、土壤等自然条件不同, 以及品种、栽培、调制技术等因素的影响, 各主产区生产的烟叶外观质量各具特点, 科学、准确地评价不同主产烟区典型生态点的烟叶外观质量, 是全国烟叶品质分析评价的重要内容之一。依据烤烟外观质量评价结果, 一方面指导烟叶生产技术措施的改进, 另一方面为卷烟工业企业的烟叶使

用提供参考依据。邓小华等^[1]研究表明, 湖南产区不同生态区域烟叶外观质量分值是湘南烟区 < 湘中烟区 < 湘西北烟区, 不同等级、不同生态区域烟叶外观质量类型差异明显。厉福强等^[2]分析了贵州湄潭烟叶的外观质量因素特征及相关性和年度稳定性。马莹等^[3]对黔西南州烟叶外观质量进行了分析。张国等^[4]运用因子分析法对湖南代表烟区烟叶

作者简介: 杨明峰, 男, 工程师, 主要从事烟叶外观质量标准研究和原烟配方打叶工作。E-mail: mingfeng618@163.com.cn

收稿日期: 2011-06-01

修回日期: 2011-12-21

外观质量指标进行了分析,胡建军等^[5]应用两阶段聚类法对云南省大理州的部分中部烟叶外观质量进行聚类分析,并应用多元方差分析对聚类效果进行了检验。然而,不同烟草种植区烟叶的外观质量区域分类未见报道。本研究运用 K - Means Cluster (快速聚类法)聚类方法分析各烟草种植区烤烟外观质量区域特征及区域分布,为卷烟工业企业合理利用不同产区烟叶原料提供数据参考。

1 材料与方法

1.1 材料

根据全国烤烟主产区生态特点,采集了云南、贵州、四川、福建等 18 个省 112 个典型生态点的 2009 年初烤 C3F 等级烟叶样品,共 112 个。每个样品取样量不少于 3 kg,品种为该典型生态点种植面积最大的主栽品种。文中对各烟草种植区不同生态点分别用代号表示。

1.2 方法

1.2.1 烤烟外观质量评价指标体系构建及烟草种植区的划分 烤烟外观质量评价指标体系的指标筛选、指标量化方法、鉴定方法、指标权重确定以及烟草种植区的划分参照中国烟草种植区划研究成果^[6]。

1.2.2 聚类分析 采用了 K-Means Cluster^[7-8]聚类方法,先给定划分的数目 k ,划分时首先创建一个初始划分,然后采用一种迭代的重定位技术,尝试通过对象在划分间的移动来改进划分,逐步调整,得到最终分类。划分结果必须同时满足以下 2 个要求:(1) 每个组至少包括 1 个对象;(2) 每个对象必须属于且只属于 1 个组。

依据生态相似优先的原则,将我国烟草种植区分为北方烟草种植区、黄淮烟草种植区、长江中上游烟草种植区、东南烟草种植区和西南烟草种植区,然后再依据烤烟外观质量特征(颜色、成熟度、叶片结构、身份、油分和色度等 6 个指标)的相似性,运用 K-Means Cluster 聚类方法分析各烟草种植区烤烟外观质量区域特征及区域分布。

1.2.3 数据处理方法 数据处理及分析在 Excel 2003、SPSS 11.0 等^[9]软件的支持下进行。

2 结果

2.1 5 大种植区中部烤烟外观质量分类及区域分布

2.1.1 北方烟草种植区 从烤烟外观质量各类对应颜色、成熟度、叶片结构、身份、油分和色度指标的分值来看(表 1、2,图 1),北方烟草种植区中部烤烟外观质量可分为 3 类。总的来看,第 3 类各指标是最优的,第 1 类次之,第 2 类一般。第 3 类烤烟外观颜色纯正,成熟度好,叶片结构疏松、富有弹性,身份适中、内含物质充实,富油分,色度强;占样品总数的 20.00%。第 1 类烤烟外观颜色

表 1 最终的类中心

Table 1 Final cluster centers

种植区	外观指标	类中心		
		1	2	3
北方	颜色	8.03	7.95	9.25
	成熟度	8.23	8.20	9.35
	叶片结构	7.62	8.25	9.00
	身份	8.10	7.85	8.85
	油分	6.88	5.00	8.25
	色度	6.40	5.85	7.70
黄淮	颜色	7.75	8.27	9.00
	成熟度	7.30	8.13	9.00
	叶片结构	6.80	8.12	9.00
	身份	7.50	7.75	9.00
	油分	5.30	6.37	6.50
	色度	5.25	6.10	6.50
长江中上游	颜色	8.40	9.00	8.65
	成熟度	8.26	8.75	8.53
	叶片结构	8.20	8.50	8.56
	身份	6.78	7.35	7.86
	油分	5.82	6.80	6.65
	色度	5.94	7.30	5.98
东南	颜色	7.83	7.92	8.51
	成熟度	7.85	8.14	8.60
	叶片结构	8.08	8.02	8.68
	身份	6.90	8.06	8.49
	油分	6.13	6.90	7.40
	色度	6.03	6.44	6.64
西南	颜色	8.52	8.16	9.30
	成熟度	8.39	8.08	9.33
	叶片结构	8.23	7.90	9.28
	身份	8.03	6.70	8.62
	油分	6.57	5.38	7.91
	色度	6.27	5.43	7.89

表2 样品类属及其与类中心的距离

Table 2 Cluster membership			
种植区	采样点	类属	距离
北方	辽宁 KY	1	1.008
	吉林 LH	1	1.477
	甘肃 ZN	1	0.983
	辽宁 KD	3	0.557
	辽宁 JP	1	0.535
	黑龙江 BX	2	0.678
	黑龙江 JX	1	1.266
	黑龙江 FJ	1	0.556
	甘肃 HX	2	0.678
	吉林 WQ	3	0.557
	河南 LY	2	1.163
	河南 JX	2	0.822
	河南 XC	2	0.565
	河南 QS	3	0.000
黄淮	山东 JN	1	0.644
	山东 YS	2	1.211
	山东 JX	2	0.824
	山东 ZC	1	0.644
	陕西 XY	2	0.985
	湖北 XF	3	0.693
	湖北 XE	3	1.007
	湖北 LC	1	0.469
	重庆 QJ	3	1.031

长江中上游	重庆 WS	2	0.791
	重庆 WL	3	0.316
	安徽 QC	1	0.851
	安徽 WH	2	1.170
	安徽 XZ	2	0.466
	福建 JY	3	1.057

	湖南 NY	3	0.572
	江西 SC	2	0.770

东南	江西 XF	2	1.153
	贵州 AL	1	1.086
	贵州 XX	1	1.154
	贵州 KY	1	0.634
	贵州 PX	3	0.759

	云南 YS	3	1.019
	云南 JC	1	1.800
	云南 ZY	1	1.632

较纯正,成熟度一般,叶片结构较疏松,身份适中,油分有,色度为“强”、“中”界限,各指标相对均衡;占样品总数的60.00%。第2类烤烟外观颜色纯正度一般,成熟度一般,叶片结构较疏松,身份适中偏厚,油分略欠,色度中;占样品总数的20.00%。

由表3看出,聚类分析的6个指标在所分的3

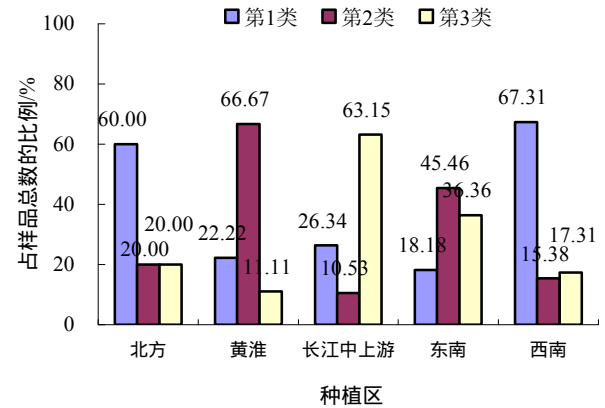


图1 烤烟外观质量分类及比例
Fig.1 Classification and proportion of appearance quality of flue-cured tobacco

类中除身份和成熟度指标外都出现了明显的差异。 F 统计量的相伴概率远小于显著性水平0.05。因此,可以认为将样本分成3类的聚类分析基本上是成功的,聚类效果比较理想。“油分”和“色度”指标在北方烟草种植区烤烟间的差异达极显著水平。

2.1.2 黄淮烟草种植区 由表1、2,图1看出,黄淮烟草种植区中部烤烟外观质量可分为3类。综合来看,烤烟外观质量第3类各指标较优,第2类次之,第1类稍差。第3类烤烟外观颜色纯正,成熟度好,叶片结构疏松、富有弹性,身份适中、内含物质充实,富油分,色度较强;占样品总数11.11%。第2类烤烟外观颜色较纯正,成熟度一般,叶片结构较疏松,身份适中,油分有,色度为“强”、“中”界限,各指标相对均衡;占样品总数的66.67%。第1类烤烟外观颜色纯正度一般,成熟度一般,叶片结构较疏松,身份适中偏厚,油分略欠,色度中;占样品总数的22.22%。

对聚类后形成的各类样本之间作单因素方差分析(表3)表明,6个指标在所分的3类中除身份指标外都出现了明显的差异,所以有5个指标在聚类分析中发挥了明显作用。从方差分析结果看, F 统计量的相伴概率远小于显著性水平0.05,因此,可以认为将样本分成3类的聚类分析基本上是成功的,聚类效果比较理想。“叶片结构”指标在黄淮烟草种植区烤烟样品间的差异达极显著水平。

表 3 各类样本方差分析检验结果
Table 3 ANOVA of various types of samples

种植区	外观指标	F 值	P
北方	颜色	6.317	0.027
	成熟度	3.737	0.079
	叶片结构	7.436	0.019
	身份	3.382	0.094
	油分	31.985	0.000
	色度	15.214	0.003
黄淮	颜色	6.949	0.027
	成熟度	5.783	0.040
	叶片结构	12.627	0.007
	身份	2.910	0.131
	油分	4.160	0.074
	色度	5.154	0.050
长江中上游	颜色	2.317	0.131
	成熟度	2.548	0.109
	叶片结构	2.975	0.080
	身份	8.681	0.003
	油分	5.893	0.012
	色度	10.614	0.001
东南	颜色	7.735	0.003
	成熟度	7.208	0.005
	叶片结构	6.244	0.008
	身份	15.807	0.000
	油分	19.313	0.000
	色度	3.817	0.040
西南	颜色	50.508	0.000
	成熟度	19.135	0.000
	叶片结构	9.417	0.000
	身份	15.042	0.000
	油分	30.593	0.000
	色度	52.091	0.000

2.1.3 长江中上游烟草种植区 由表 1、2，图 1 看出，长江中上游烟草种植区中部烤烟外观质量可分为 3 类，综合来看，烤烟外观质量第 2 类各指标较优，第 3 类次之，第 1 类一般。第 2 类烤烟外观颜色纯正，成熟度好，叶片结构疏松、富有弹性，身份适中略欠，富油分，色度强；占样品总数的 10.53%。第 3 类烤烟外观颜色较纯正，成熟度较好，叶片结构较疏松，身份适中，油分有，色度为“强”、“中”界限，各指标相对均衡，占样品总数的 63.16%。第 1 类烤烟外观颜色纯正度稍好，成熟度一般，叶片结构较疏松，身份偏薄，油分略欠，色度中；占样品总数的 26.32%。

由表 3 看出，聚类分析的 6 个指标在所分的 3 类中身份、油分和色度出现了明显的差异，所以有

3 个指标在聚类分析中发挥了明显作用。从方差分析结果看， F 统计量的相伴概率远小于显著性水平 0.05，因此，可以认为将样本分成 3 类的聚类分析基本上是成功的，聚类效果比较理想。“身份”和“色度”指标在长江中上游烟草种植区烤烟样品间的差异达极显著水平。

2.1.4 东南烟草种植区 由表 1、2，图 1 看出，东南烟草种植区中部烤烟外观质量可分为 3 类，综合来看，烤烟外观质量第 3 类各指标较优，第 2 类次之，第 1 类一般。第 3 类烤烟外观颜色较纯正，成熟度较好，叶片结构疏松、富有弹性，身份适中，富油分，色度强；占样品总数的 36.36%。第 2 类烤烟外观颜色较纯正，成熟度一般，叶片结构较疏松，身份适中，油分有，色度为“强”、“中”界限，各指标相对均衡；占样品总数的 45.45%。第 1 类烤烟外观颜色纯正度稍好，成熟度一般，叶片结构较疏松，身份偏薄，油分略欠，色度为“强”、“中”界限；占样品总数的 18.18%。

由表 3 看出，聚类分析的 6 个指标在所分的 3 类中均出现了明显的差异，所以 6 个指标在聚类分析中都发挥了明显作用。从方差分析结果看， F 统计量的相伴概率远小于显著性水平 0.05，因此，可以认为将样本分成 3 类的聚类分析是成功的，聚类效果比较理想。“身份”和“油分”指标在东南烟草种植区烤烟样品间的差异达极显著水平。

2.1.5 西南烟草种植区 由表 1、2，图 1 看出，西南烟草种植区中部烤烟外观质量可分为 3 类。综合来看，烤烟外观质量第 3 类各指标是最优的，第 1 类次之，第 2 类一般。第 3 类烤烟外观颜色纯正，成熟度好，叶片结构疏松、富有弹性，身份适中、内含物质充实，富油分，色度强；占样品总数的 17.31%。第 1 类烤烟外观颜色较纯正，成熟度较好，叶片结构较疏松，身份适中，油分有，色度为“强”、“中”界限，各指标相对均衡；占样品总数的 67.31%。第 2 类烤烟外观颜色纯正度一般，成熟度一般，叶片结构较疏松，身份适中偏薄，油分略欠，色度中；占样品总数的 15.38%。

从方差分析结果看(表3), F 统计量的相伴概率远小于显著性水平 0.01, 因此, 可以认为将样本分成 3 类的聚类分析是成功的。聚类效果比较理想。烤烟外观各项指标在西南烟草种植区烤烟样品间差异均达极显著水平。

为了检验聚类分析结果, 对聚类分析后形成的各类样本之间作单因素方差分析(表3)。聚类分析的 6 个指标在所分的 3 类中均出现了明显的差异, 所以 6 个指标在聚类分析中都发挥了明显作用。

3 小 结

5 大烟草种植区中部烟叶烤烟样品间均存在明显差异, 不同种植区指标差异不同。

北方烟草种植区烤烟样品“油分”和“色度”指标存在极显著差异, 烟叶颜色、成熟度和身份一致性较好, 但油分不足。

黄淮烟草种植区烤烟样品“叶片结构”指标存在极显著差异, 多数烤烟成熟度一般, 叶片结构趋紧, 应引起重视。

长江中上游烟草种植区烤烟样品“身份”和“色度”指标存在极显著差异, 多数烤烟内含物略欠充实, 身份偏薄。

东南烟草种植区烤烟样品“身份”和“油分”指标存在极显著差异, 多数烤烟内含物略欠充实,

身份偏薄, 油分略欠。

西南烟草种植区烤烟样品烤烟外观各指标均存在极显著差异, 可能与西南烟区立体气候差异有关, 多数烤烟颜色、成熟度、叶片结构较好, 但油分和色度差异较大。

参考文献

- [1] 邓小华, 周冀衡, 杨虹琦, 等. 湖南烤烟外观质量量化评价体系的构建与实证分析[J]. 中国农业科学, 2007, 40(9): 2036-2044.
- [2] 厉福强, 张长云, 张长华. 贵州湄潭烤烟外观质量因素分析[J]. 贵州农业科学, 2009, 37(1): 54-57.
- [3] 马莹, 朱云燕, 田野, 等. 黔西南州烟叶外观质量分析[J]. 中国农学通报, 2009, 25(4): 78-81.
- [4] 张国, 王奎武, 朱列书, 等. 湖南烤烟外观质量因子分析[J]. 中国农学通报, 2007, 23(2): 114-116.
- [5] 胡建军, 周冀衡, 张建平, 等. 两阶段聚类分析在烤烟外观质量评价中的应用[J]. 农业机械学报, 2009, 40(6): 143-146.
- [6] 王彦亭, 谢剑平, 李志宏, 等. 中国烟草种植区划图册[M]. 西安: 西安地图出版社, 2008: 1-5.
- [7] Tang Q Y, Feng M G. DPS Data processing system: Experimental design, statistical analysis, and data mining [M]. Beijing: Science Press, 2007.
- [8] 高惠璇. 应用多元统计分析[M]. 北京: 北京大学出版社, 2005.
- [9] 郝黎仁. SPSS 实用统计分析[M]. 北京: 中国水利水电出版社, 2002.