

DOI:10.6041/j.issn.1000-1298.2012.12.024

西湖龙井茶品质的智能嗅觉识别*

史波林¹ 赵 镭¹ 支瑞聪¹ 席兴军¹ 朱大洲²

(1. 中国标准化研究院食品与农业标准化研究所, 北京 100088; 2. 国家农业智能装备工程技术研究中心, 北京 100097)

【摘要】 通过电子鼻采集不同等级、树种和产地西湖龙井茶的智能嗅觉指纹图谱,利用主成分分析得分矩阵研究等级、树种和产地指纹信息对茶叶品质的影响程度,基于软独立模型分类分析方法建立茶叶等级、树种和产地的3类智能嗅觉判别模型。结果表明,不同等级西湖龙井茶的电子鼻信号差异最大;在涵盖不同树种和产地信息的样品中,电子鼻能准确预测品质相近的高档等级(精品、特级和一级)茶叶,等级判别正确率基本达到100%。树种与产地特征对于茶叶品质的影响程度比较接近,并且同一等级、同一产地不同树种模型和同一等级、同一树种不同产地模型的判别正确率基本都达到92%以上。在此基础上,提出了首先利用电子鼻进行等级划分,然后在同一等级下进行树种鉴定和产地判别的西湖龙井茶品质智能嗅觉快速检测策略。

关键词: 西湖龙井茶 电子鼻 智能嗅觉 特征识别

中图分类号: TP242.6⁺4; TS272.7 **文献标识码:** A **文章编号:** 1000-1298(2012)12-0130-06

Quality Recognition of Xihu Longjing Tea Based on Intelligent Olfactory

Shi Bolin¹ Zhao Lei¹ Zhi Ruicong¹ Xi Xingjun¹ Zhu Dazhou²

(1. Food and Agriculture Standardization Institute, China National Institute of Standardization, Beijing 100088, China

2. National Research Center of Intelligent Equipment for Agriculture, Beijing 100097, China)

Abstract

The intelligent olfactory spectrograms of Xihu Longjing tea were collected by electronic nose. The spectrograms scores matrix that obtained by principal component analysis was applied. The influence of the information from grading, tree varieties and producing areas on quality of tea was analyzed. The soft independent modeling of class analogy was used to establish three kinds of models for grading, tree varieties and producing areas, respectively. Results showed that the intelligent olfactory spectrograms from different grading tea had the most difference than the other two elements. The grading discrimination model had well performance with about 100% correct rate to predict top grade teas. Meanwhile, the influences of tree varieties and location on the tea quality were seemed to be similar. Under the same grade and same producing area, the tree varieties discrimination model of tea was built. Also, the producing area discrimination model of tea was built. The correct rates of the two models were all over 92%. On the basis of above study, the quick detection strategy of Xihu Longjing tea by intelligent olfactory technology was presented, that is, grading demarcation, tree varieties identification and producing area distinguishing under the same grading were carried out in turn.

Key words Xihu Longjing tea, Electronic nose, Intelligent olfactory, Characteristics recognition

收稿日期: 2011-11-25 修回日期: 2012-05-22

* 国家高技术研究发展计划(863计划)资助项目(2011AA1008047)和北京市优秀人才培养资助项目(2012D00999900001)

作者简介: 史波林, 副研究员, 主要从事食品感官评价及智能感官分析研究, E-mail: spring0214@163.com

通讯作者: 赵镭, 副研究员, 主要从事食品感官标准化研究, E-mail: zhaolei@cnis.gov.cn

引言

西湖龙井市场以次充好现象盛行^[1]。同时随着市场的不断扩大,有些角落也没能彻底杜绝龙井茶假冒伪劣问题,使得监管存在一定困难^[2]。龙井群体和龙井 43 号这两类树种是当地的主要种植茶树^[3]。为了提升西湖龙井茶的品质,学者提出进一步细分以产地为特色的西湖龙井茶^[4]。这对龙井茶特定西湖产区内部的小产地鉴定提出了更高的要求。

电子鼻分析技术具有检测时间短、重复性好、无需复杂样品预处理过程、不发生感官疲劳和检测结果客观可靠等特点,更重要的是可以在一定程度上模拟人的感官给出有关茶叶整体香气的评判结果和指纹信息^[5]。目前已有电子鼻研究茶叶品质的相关工作^[6-12]。这些研究主要以差异性大的茶叶样品为研究对象,而本文以小范围内、品质差异小的西湖龙井茶为对象,对定性判别研究提出更高要求。利用电子鼻的茶叶香气响应图谱分析等级、产地和树种等 3 个因素对茶叶品质差异的影响程度,基于模式识别方法结合不同茶树所引起的植物生长特点分析其品质判别模型的性能差异根源,建立西湖龙井茶等级划分、树种鉴定和产地判别的快速检测方法。

1 实验材料与方法

1.1 茶样制备

收集来自 2010 年西湖产区不同等级、不同树种、不同产地的 8 种西湖龙井茶,并进行编号(3 个号码从左到右分别依次代表树种、产地和等级信息),如表 1 所示。采用密封性能良好的铝箔袋(北京华盾塑料有限公司,10 cm × 10 cm,食品级,无毒无味)将茶样独立真空包装成 3 g/袋,并置于 -4℃ 以下的冷库内保存^[13]。按照实验用量每次取若干小袋即可。

表 1 西湖龙井茶样品信息

Tab. 1 Information of Xihu Longjing tea samples

序号	树种	产地	等级	编号	样品数
1	龙井 43 号	翁家山	精品	LWJ	49
2	龙井 43 号	杨梅岭	精品	LYJ	33
3	群体	翁家山	精品	QWJ	40
4	群体	虎跑后山	精品	QHJ	50
5	群体	龙坞	精品	QLJ	44
6	龙井 43 号	梅家坞	特级	LMT	44
7	群体	梅家坞	特级	QMT	45
8	龙井 43 号	翁家山	一级	LW1	85

1.2 智能嗅觉采集方法

采用法国 Alpha MOS 公司生产的 Fox 4000 型传感器阵列电子鼻。该设备配置 HS100 型顶空自动系统和智能嗅觉数据采集系统,具备系统误差自诊断和校准功能。实验按照茶汤香气顶空产生、抽取顶空气体进样、传感器阵列感应、清洗顶空针与传感器、检测下一个样品的流程操作。具体方法如下:在每个 20 mL 顶空瓶中装入 1.00 g 龙井茶干茶,并加入 5 mL 常温超纯水后压盖密封;当顶空瓶被送入预热区时,在 500 r/min 振荡器转速和 60℃ 顶空温度下加热 900 s 后,抽出 2.0 mL 气体以 2.0 mL/s 的进样速度注入到电子鼻传感器阵列室(含 18 根金属氧化物传感器)。在 120 s 响应时间内,将所引起的传感器相对电阻 S 作为样品的气味指纹图谱(图 1)。 $S = (R - R_0)/R_0 = \Delta R/R_0$,其中 R 为传感器在待测气体中的电阻, R_0 为传感器在合成干燥空气中的电阻(即传感器空载时的电阻)。选择传感器响应绝对值的最大点,即气味指纹图谱的波峰点或波谷点,作为后续茶叶品质模型建立的特征点。

茶样通过循环交叉顺次排列检测,以避免设备所产生的系统误差。另外,每次进样针清洗时间为 120 s,传感器清洗时间为 300 s。

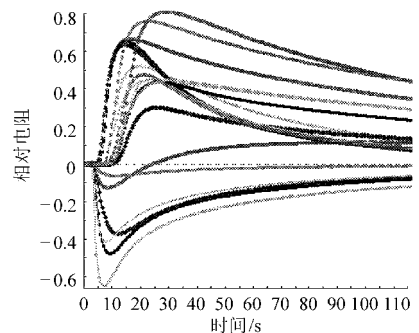


图 1 龙井茶的传感器响应信号强度图

Fig. 1 Sensor signal intensity of Longjing tea

1.3 数学统计方法

采用主成分分析(PCA)进行数据降维^[14],通过主成分得分反映茶样间的相似性和独特性,揭示样本的内部特征和聚类信息,进一步说明每个样本在大类样本集中是否存在较大差异,是否根据样品间品质特征形成自动聚类现象。

采用软独立模型分类分析(SIMCA)^[15]建立西湖龙井茶不同等级、不同树种及不同产地的分类识别模型。此方法以每类样品的 PCA 得分为基础,以 SIMCA 距离为依据,注重各类样品内部相似程度为建模特征。在模型研究中,将样本集划分为校正集和预测集,采用校正集样品建立定性判别分类模型,通过预测集样品考察所建模型的准确性和可靠性。

在此,随机选择总样品的 2/3 作为校正集样品,剩下的 1/3 用作预测集样品,使得校正集既具有较好的代表性,同时又拓宽模型的预测范围,增强模型的适应能力^[16],并且以正确率、识别率和拒绝率作为模型预测性能指标。正确率代表样品被准确判别各自类别(等级或树种或产地)的样品数之和与所有被测样品数的比值;识别率为各类别样品被准确分类的样品数与本类别样品总数的比值;拒绝率为其他类别样品没有被误判为本类别样品数与非本类别样品总数的比值。所有算法利用 Matlab 7.0 自行编程实现。

2 结果与分析

2.1 等级、树种、产地因素对品质的影响

西湖龙井茶品质影响因素众多,在规范化的现代茶叶加工水平下,对于同一年份的茶叶,主要考虑其等级、树种及产地因素特征。

对于涵盖等级、树种和产地信息的所有 390 个西湖龙井茶样品智能嗅觉特征响应点的 PCA 结果(图 2a~2c)可知,PC1 与 PC2 的贡献率分别为 94% 与 4%,即前 2 个主成分累计贡献率高达 98%,远超过统计学一般认定 85% 的界限^[17]。充分说明通过 PCA 分析,可以将以 18 根传感器为变量的智能嗅觉图谱矩阵转换为以 PC1 和 PC2 为变量的得分矩阵,并保留了西湖龙井茶香气品质智能图谱 98% 的信

息,完全可以此为基础进行分析。在 PCA 得分图中,具有相同或相近性质的茶样聚集在一起,而差异较明显的茶样相互远离。图 2a、2b、2c 分别以等级、树种和产地进行编号,以探索这 3 类因素对茶样 PCA 得分聚类的影响。对比这 3 个图,涵盖等级、树种和产地信息的所有样品 PCA 得分呈现明显的等级聚类特征,即在 PC1 方向上,从左到右依次为一级、特级和精品顺序排列,由此说明通过 PCA 处理后,PC1 代表了等级分类的信息,并且等级对于西湖龙井茶的品质影响程度远超过了树种和产地的影响。此结果与等级作为商品茶定质论价唯一依据相符合。

为了避免等级信息的干扰,单独对精品龙井茶进行 PCA 分析,以探究同一等级下的树种和产地信息分布特征。图 2d 为精品西湖龙井茶样在不同产地、不同树种下的 PCA 得分图,由图可知虽然 PC1 与 PC2 的累计贡献率达到 87% (>85%),已基本代表了样品智能嗅觉的整体信息,但 PC1 的贡献率(56%)与 PC2 的贡献率(31%)相差不多,PC1 未达到主导茶叶某类品质特征的贡献,由此推断图 2d 中的 PC1 可能涵盖了代表样品品质的多种信息。因此,在图中体现不出明显的按照树种信息聚类或按照产地信息聚类的现象,初步推断树种和产地因素可能对西湖龙井茶品质的影响程度比较接近。

为了探讨树种和产地信息的特征,需要在同一

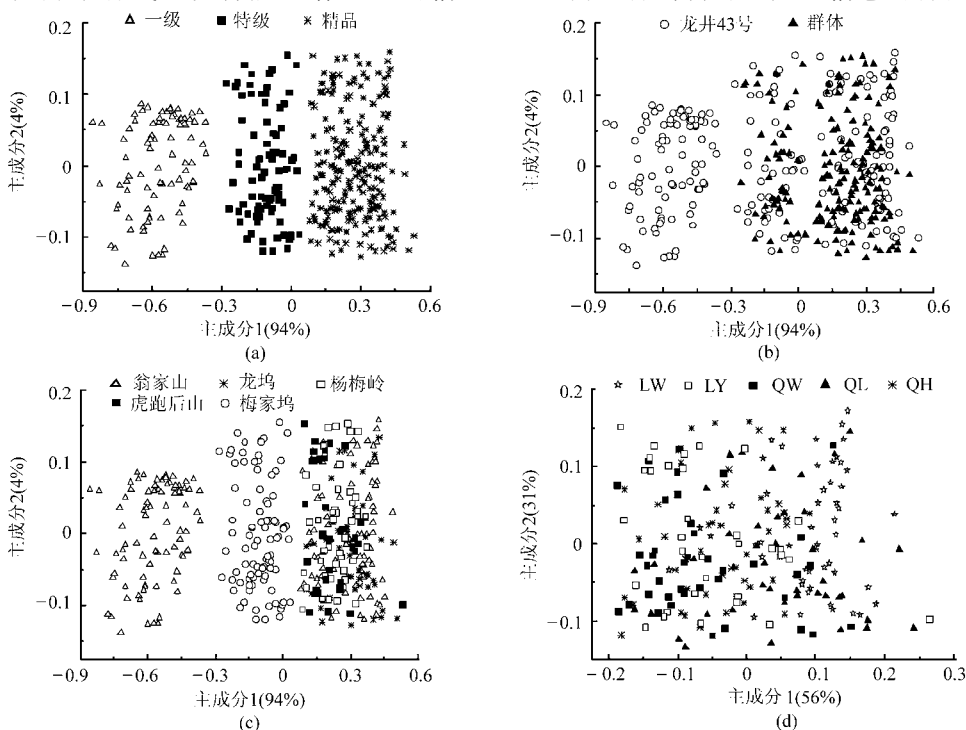


图 2 西湖龙井茶叶 PC1-PC2 得分图

Fig. 2 Score plot using PC1-PC2 for Xihu Longjing tea samples

(a) 按等级编号的全体茶样 (b) 按树种编号的全体茶样 (c) 按产地编号的全体茶样 (d) 所有精品茶样

等级同一产地条件下研究树种模型以及在同一等级同一树种条件下研究产地模型。因此基于当初所采集的 8 种茶样,树种模型有 2 个,分别为 LMT-QMT 和 LWJ-QWJ;产地模型也有 2 个,分别为 LWJ-LYJ

LYJ 和 QLJ-QWJ。分别对此 4 组茶样集进行 PCA 分析,各模型内样品间能形成各自类别区域,其中 LMT 与 QMT 在 PC5 上区分非常明显,其他 3 个模型在 PC1 上有明显区分,如图 3 所示。

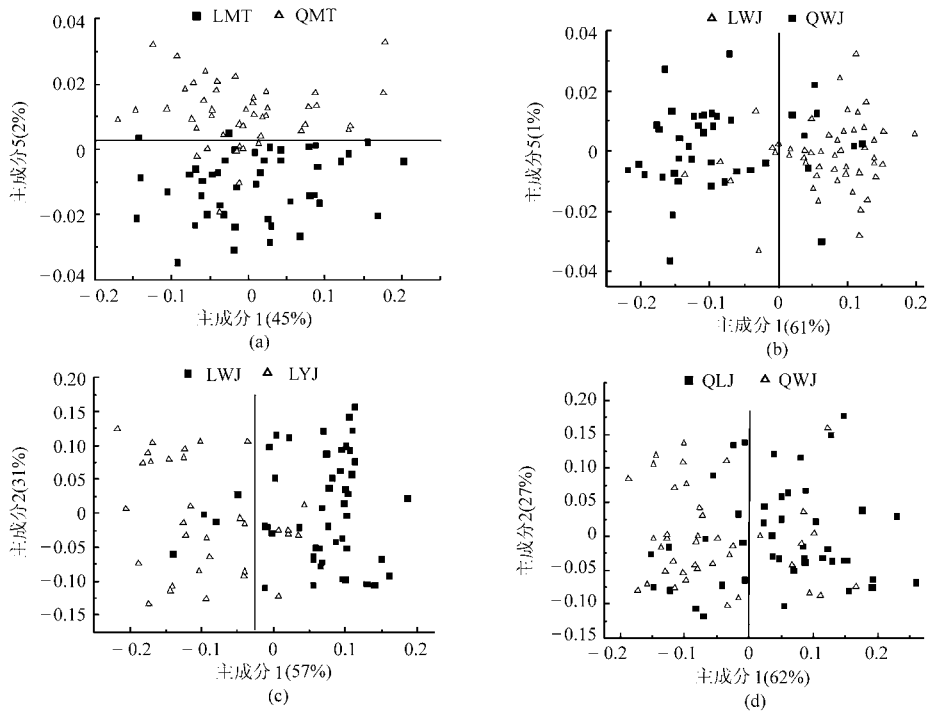


图 3 西湖龙井茶叶主成分分析得分图

Fig. 3 Score plot using PCA for Xihu Longjing tea samples

(a) LMT-QMT (b) LWJ-QWJ (c) LWJ-LYJ (d) QLJ-QWJ

2.2 等级品质识别

为了满足研究要求,尽量包含树种差异和产地差异的样品,同时避免因不同等级样品数量不均,导致分类模型的不平衡,即引起对建模样品数量多的等级预测性能高,而对建模样品数量少的等级预测性能低。因此,从 8 种所采集茶样中(表 1)挑选出 5 种样品用于等级建模研究,并同样按照 2:1 的比例将样品随机划分成校正集和预测集(表 2)。

表 2 等级模型的不同样本集数量

Tab. 2 Calibration and prediction samples numbers of grading model

样本等级	校正集	预测集	总样本数
精品 (LYJ, QHJ)	55	28	83
特级 (LMT, QMT)	60	29	89
一级 (LW1)	56	29	85
所有样本数	171	86	257

SIMCA 算法是基于 PCA 类模型的方法,确定最佳主成分数是模型建立的关键环节之一。在建模中,若所取的主成分数过少,则因关于等级分类的信息不足,使得模型性能不佳;随着主成分数的增加,等级信号增加,模型预测性能也提高;但主成分数超

过一定数值时,随着等级信号增加,噪声也大幅提高,从而出现过拟合现象,导致模型预测性能又降低。本研究通过交互验证法来确定不同等级茶叶模型的最佳主成分数,即在预测残差平方和 (PRESS) 最小值处选取主成分数。图 4 是不同等级样品模型的 PRESS 与主成分数之间的关系。由图可知,随着主成分数的增加, PRESS 值先降低,然后再提高。精品、特级和一级的最佳主成分数在 6、7、8 时,各自对应的 PRESS 最小,也就是此时的主成分数为各类建模所需要的最佳值。

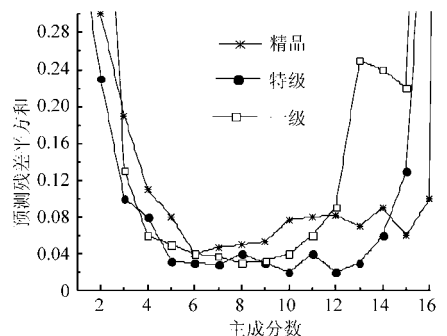


图 4 预测残差平方和与模型主成分数之间的关系
Fig. 4 Relationship between prediction residual error sum of squares and principal components in model

在不同等级类样品的最佳主成分数下建立 SIMCA 等级判别模型,表 3 为建模结果。其校正集各等级样品的识别率、拒绝率和等级分类正确率都为 100%,也就是没有误判现象。预测集中,只有 1 个特级样品误判为一级样品,其他都准确预测。由此可知,通过电子鼻的智能嗅觉技术结合 SIMCA 建模方法,能有效判别品质差异非常小的不同等级高档西湖龙井茶,并不受树种和产地的影响,完全满足检测需要。

表 3 茶叶等级模型的判别结果

Tab.3 Discriminate analysis results of grading model for tea

参数		精品	特级	一级
校正集	样品数	55	60	56
	识别率/%	100	100	100
	拒绝率/%	100	100	100
	正确率/%		100	
预测集	样品数	28	29	29
	识别率/%	100	96.65	100
	拒绝率/%	100	100	98.25
	正确率/%		98.84	

2.3 树种及产地品质识别

由于同一类样品来源于多家当地茶叶农户,其不同的制作手法、经验和炒制温度的控制能力,使得同类样品间的品质会有一定程度的差异,因此在图 3 中也有一小部分样品重叠的现象。并且 PCA 算法本身更加注重数据矩阵空间坐标的转换和样品信息的整体性保留,只是在一定程度上体现样品品质间的内聚性和差异性。为了进一步分析树种与产地因素对样品品质的影响,以及智能嗅觉检测数学建模对不同树种与不同产地的区分能力和区分程度,在此采用 SIMCA 定性分类判别方法。

SIMCA 算法可体现各类别茶样内部的相似程度,它首先对每一类样品进行 PCA 建模,使同类样本聚集在同一空间区域。其中各主成分代表同类样品的变异特性,越靠前的主成分包含的树种或产地特征越丰富,对分类的作用越大,在此也采用交互验证法来确定不同类样品的最佳主成分数。表 4 为同一等级、同一产地下的树种 SIMCA 建模及同一等级、同一树种下的产地 SIMCA 建模结果。

结果显示,各种模型中不同类样品的最佳主成分数都为 6 或 7。同时可知,虽然每个模型前 2 个主成分的累计贡献率都能超过 85%,代表绝大部分茶样品品质信息,但 PC3 以后的信息对建模也起到非常大的作用,这些主成分虽然不包含体现茶叶品质的主体及不同类间的共性信息,但却包含了能体现样品间差异的信息。比如 LMT-QMT 模型中 PC5 虽然贡献率只有 2%,但在主成分得分图中却能很好地区分 2 类树种样品。表 4 显示,校正集和预测集的样品正确率比较接近,说明建模用的校正集样本具有代表性,所建模型稳定可靠。同时这些模型的分正确率都超过 92% (除 QLJ-QWJ 的产地模型),预测性能良好。QLJ-QWJ 模型性能虽然低于其他 3 个模型,但也在 90% 附近,这一趋势与其对应的主成分得分图相吻合。其主要原因是:相对于改良后的龙井 43 号树种,传统的群体树种由于是长期的自然杂交和有性繁殖,个体间在发芽期、长势、芽叶大小和色泽、生化成分含量等方面都有着不同程度的差异,使其在同一茶园中同批采摘的群体嫩叶制成的茶叶香气也有一定程度的差异。也正是这个原因,树种模型的整体性能好于产地模型;而在产地模型中,同样龙井 43 号树种下的模型效果要好于都是群体品种下的不同产地判别性能。

表 4 树种模型和产地模型的判别结果

Tab.4 Discriminate analysis results of tree varieties models and producing areas models

模型类别		校正集			预测集		
		正确数	错误数	正确率/%	正确数	错误数	正确率/%
树种模型	LMT(6)-QMT(7)	56	3	94.92	28	2	93.33
	LWJ(7)-QWJ(7)	55	4	93.22	28	2	93.33
产地模型	LWJ(7)-LYJ(6)	52	3	94.55	25	2	92.59
	QLJ(7)-QWJ(7)	52	4	92.86	25	3	89.29

注:模型括号中的数字为各类样品 SIMCA 模型的最佳主成分数。

3 讨论

总体来讲,在结合智能嗅觉技术的西湖龙井茶品质判别中,可以由易到难分成 3 个步骤进行检测。首先通过电子鼻区分不同等级;根据其他学者的研

究结果,来源于不同树种的茶叶制品在外观上差异较大,可以通过视觉分析判别茶叶的不同树种来源^[18],因此在等级区分后可以进行树种判别;结合前 2 个环节,就可以用电子鼻判别茶叶西湖龙井产区下的不同小区域产地,也就是分析同一等级同一

树种下的西湖龙井茶叶产地来源。由此可以达到等级、树种和产地判别的目的,为西湖龙井茶的品质界定与特征提炼提供基础,为茶叶的品质检测和掺假鉴别提供技术支撑。

4 结束语

西湖龙井茶品质的等级特征体现最为明显,在不考虑其他因素的前提下,可以通过智能嗅觉建模直接识别茶叶的等级,特别对于品质相近的高档次样品,其预测效果较好,校正集各等级判别正确率为

100%,预测集中只有 1 个特级样品被误判为一级样品。树种和产地特征对于西湖龙井茶叶品质的影响程度比较接近,然而在同一等级、同一产地下的树种判别,以及在同一等级、同一树种下的产地判别,都具有良好的模型性能,不同类别样品判别正确率都超过 92%。由于群体树种嫩叶炒制的样品个体差异性大,有助于同一等级、同一产地下的不同树种判别,使得树种模型的整体性能好于产地模型;也使得同一等级、同一龙井 43 号树种下的不同产地判别效果要好于同一等级、同一群体树种下的产地模型。

参 考 文 献

- 1 李晓丽,何勇. 基于多光谱图像及组合特征分析的茶叶等级区分[J]. 农业机械学报, 2009, 40(增刊): 113~118.
Li Xiaoli, He Yong. Classification of tea grades by multi-spectral images and combined features[J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(Supp.): 113~118. (in Chinese)
- 2 王玉玲. 西湖龙井茶市场失信的经济学研究[D]. 杭州: 浙江大学, 2008.
Wang Yuling. Economic research on the dishonesty of the west lake dragon tea market[D]. Hangzhou: Zhejiang University, 2008. (in Chinese)
- 3 DB3301/T 005.4—2009 杭州市农业标准规范 西湖龙井茶 第 2 部分 栽培技术[S]. 2009.
- 4 郑小洁,连勇,郑娟婷. 关于完善“西湖龙井茶”地方标准的设想[J]. 杭州, 2010(3): 35.
- 5 Burattia S, Benedetti S, Scampicchio M. Characterization and classification of Italian Barbera wines by using an electronic nose and an amperometric electronic tongue[J]. Analytica Chimica Acta, 2004, 525(1): 133~139.
- 6 Katayama S, Takeuchi, Taniguchi H, et al. Odor analysis of green tea by sensors[C]// Proceedings of 2004 International Conference-CHA(tea) Culture and Science, 2004: 642~646.
- 7 Chen Lisue, Ou Andi Shau-mei. Comparison of Taiwan tea aroma analysis by sensory analysis and electronic nose[C]// Proceedings of 2005 International Symposium on Innovation in Tea Science and Sustainable Development in Tea Industry, 2005: 629~638.
- 8 Ritaban Dutta E L, Hines J W, Gardner K R, et al. Tea quality prediction using a tin oxide-based electronic nose: an artificial intelligence approach[J]. Sensors and Actuators B, 2003, 94(2): 228~237.
- 9 Yu H, Wang J. Discrimination of Longjing green-tea grade by electronic nose[J]. Sensors and Actuators B, 2007, 122(1): 134~140.
- 10 Borah S, Hines E L, Leeson M S, et al. Neural network based electronic nose for classification of tea aroma[J]. Sensing and Instrumentation for Food Quality and Safety, 2008, 2(1): 7~14.
- 11 Nabaran B, Sohan S, Bipan T, et al. Monitoring of black tea fermentation process using electronic nose[J]. Journal of Food Engineering, 2007, 80(4): 1146~1156.
- 12 Nabaran B, Sohan S, Bipan T, et al. Detection of optimum fermentation time for black tea manufacturing using electronic nose[J]. Sensors and Actuators B, 2007, 122(2): 627~634.
- 13 于慧春,王俊,张红梅,等. 龙井茶叶品质的电子鼻检测方法[J]. 农业机械学报, 2007, 38(7): 103~106.
Yu Huichun, Wang Jun, Zhang Hongmei, et al. Measurement of the Longjing tea quality by using an electronic nose[J]. Transactions of the Chinese Society for Agricultural Machinery, 2007, 38(7): 103~106. (in Chinese)
- 14 Peter M F, James B L, Paul R O, et al. Materials characterization using factor analysis of FT-IR spectra. Part 2: mathematical and statistical considerations[J]. Applied Spectroscopy, 1985, 39(2): 311~316.
- 15 Chen Q, Zhao J, Zhang H, et al. Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration[J]. Analytica Chimica Acta, 2006, 572(1): 77~84.
- 16 Liu Y, Ying Y, Jiang H. Rapid determination of maturity in apple using outlier detection and calibration model optimization [J]. Transactions of ASABE, 2006, 49(1): 91~95.
- 17 Philip R C N, Paul A T, John F M. Missing data methods in PCA and PLS: score calculations with incomplete observations [J]. Chemometrics and Intelligent Laboratory Systems, 1996, 35(1): 45~65.
- 18 Wu D, Yang H, Chen X, et al. Application of image texture for the sorting of tea categories using multi-spectral imaging technique and support vector machine[J]. Journal of Food Engineering, 2008, 88(4): 474~483.