

电子鼻检测黄山毛峰茶贮藏时间方法研究

薛大为¹, 杨春兰¹, 孔慧芳², 鲍俊宏¹

(1. 蚌埠学院电子与电气工程系, 安徽蚌埠 233030)(2. 合肥工业大学电气与自动化工程学院, 安徽合肥 230009)

摘要: 利用电子鼻对6个不同贮藏时间下5个等级黄山毛峰茶进行检测。首先获取反映茶叶香气的原始特征向量, 再通过主成分分析法(PCA)提取出前5个主成分作为主特征向量, 然后以主特征向量作为BP神经网络(BPNN)的输入, 建立黄山毛峰茶贮藏时间预测模型(称为PCA-BPNN)。通过对75个测试样本(每等级15个)实验测试表明: PCA-BPNN对于贮藏0d的茶叶, 最大预测误差为7d, 5个(6.67%)样本预测误差超过10d; 对于贮藏60d的茶叶, 最大预测误差为10d, 4个(5.33%)样本预测误差超过10d; 对于贮藏120d的茶叶, 最大预测误差为16d, 7个(9.33%)样本预测误差超过10d; 对于贮藏180d的茶叶, 最大预测误差为19d, 8个(10.67%)样本预测误差超过10d; 对于贮藏240d的茶叶, 最大预测误差为21d, 8个(10.67%)样本预测误差超过10d; 对于贮藏300d的茶叶, 最大预测误差为14d, 6个(8.00%)样本预测误差超过10d。验证了PCA-BPNN预测模型用于检测黄山毛峰茶贮藏时间的可行性, 同时与以原始特征变量作为输入的BPNN预测模型相比, 性能更好。

关键词: 黄山毛峰茶; 电子鼻; PCA; BPNN; 预测模型

文章编号: 1673-9078(2016)8-328-333

DOI: 10.13982/j.mfst.1673-9078.2016.11.049

An Electronic Nose-based Method for Determination of the Storage Time of Huangshan Maofeng Tea

XUE Da-wei¹, YANG Chun-lan¹, KONG Hui-fang², BAO Jun-hong¹

(1. Department of Electronic and Electrical Engineering, Bengbu University, Bengbu 233030, China)

(2. School of Electrical and Automation Engineering, Hefei University of Technology, Hefei 230009, China)

Abstract: Five grades of Huangshan Maofeng tea samples with six different storage times were analyzed by using an electronic nose. First, the original feature vectors representing the tea odor were acquired, and the first five principal components were extracted as the principal feature vectors. The principal feature vectors were used as the input of back propagation neural network (BPNN) to establish the prediction model for the storage time of Huangshan Maofeng tea (called PCA-BPNN). The test was carried out on 75 tea samples (15 samples of every grade). The results showed that for the tea at zero day of storage, the maximum prediction error was seven days and the prediction error of five samples exceeded ten days (6.67%). For the tea of 60 d of storage, the maximum prediction error was ten days, and the prediction error of four samples exceeded ten days (5.33%). For the tea of 120 d of storage, the maximum prediction error was 16 d and the prediction error of seven samples exceeded ten days (9.33%). For the tea of 180 d of storage, the maximum prediction error was 19 d and the prediction error of eight samples exceeded ten days (10.67%). For the tea of 240 d of storage, the maximum prediction error was 21 d and the prediction error of eight samples exceeded ten days (10.67%). For the tea of 300 d of storage, the maximum prediction error was 14 d and the prediction error of six samples exceeded ten days (8.00%). The feasibility of PCA-BPNN prediction model to determine the storage time of Huangshan Maofeng tea was verified. Moreover, the performance of PCA-BPNN prediction model was better than that of BPNN prediction model using the original feature vectors as the input.

Key words: Huangshan Maofeng tea; electronic nose; principal component analysis; back propagation neural network; prediction model

收稿日期: 2015-12-12

基金项目: 安徽省高等学校省级自然科学基金项目(KJ2013Z195); 安徽省高等学校优秀青年人才基金项目(2012SQRL218); 国家级大学生创新创业训练计划项目(201511305023)

作者简介: 薛大为(1978-), 男, 副教授, 主要从事仪器仪表与智能检测、模式识别

通讯作者: 杨春兰(1980-), 女, 讲师, 主要从事仪器仪表与智能检测、食品品质检测

茶叶香气与茶叶品质密切相关,是反映茶叶品质的重要因素。然而茶叶香气是由上百种挥发性物质共同作用而形成的整体信息反映^[1],构成复杂,少数几个传感器难以对其进行有效检测。电子鼻技术的出现为检测复杂气味提供一种有效的途径。电子鼻^[2]是一种仿生物嗅觉功能的检测仪器,主要由传感器阵列和模式识别方法组成,可以对气味的整体信息进行识别。电子鼻技术在水果^[3-4]、酒类^[5-6]、肉类^[7-8]等食品工业领域已经有着广泛的应用。根据文献电子鼻技术在茶叶品质检测方面也有一定报道,但大都集中于对不同等级茶叶品质检测^[9-10],而对于不同贮藏时间茶叶品质检测方面研究报道很少。众所周知,茶叶的品质不仅取决于茶叶的等级,而且与茶叶贮藏时间的长短有关,等级越高品质越好,而贮藏时间越长品质越差。

黄山毛峰茶因盛产于安徽黄山地区而得名,是中国的名茶之一,但对其品质检测的研究报道不多,尤其是对于不同贮藏时间品质检测方面的研究报道还没有。因此,本文拟利用电子鼻技术对不同贮藏时间的黄山毛峰茶品质进行检测,为预测茶叶贮藏时间提供有效的方法。模式识别是电子鼻中重要的组成部分,BP神经网络^[11](Back propagation neural network, BPNN)具有非线性映射能力强、结构简单与泛化能力强等优点,是模式识别中广泛使用的方法。由于茶叶香气构成复杂,为了全面的获取茶叶的香气信息,通常需要选择较多的特征变量来构成原始特征向量,而这些变量之间往往会存在一定的相关性,如果以这些变量作为神经网络的输入,不但建模的效率低且准确度也会下降。因此,采用主成分分析法(Principal component analysis, PCA)与BPNN结合(以下简称为PCA-BPNN)建立黄山毛峰茶贮藏时间的预测模型。首先,通过电子鼻传感器阵列获取各等级不同贮藏时间茶叶香气的原始特征向量,再通过主成分分析法提取出能够反映香气整体信息的主特征向量,最后以主特征向量作为神经网络的输入建立茶叶贮藏时间预测模型。

1 材料与方法

1.1 茶叶样品

试验用茶叶样品是从某黄山毛峰茶厂专门订购的干茶叶。为了保证试验结果的可靠性,要求各等级新鲜茶叶均采自同一茶园且加工工艺统一,这样可以尽量保持客观条件一致,以减少随机因素的影响。茶叶样品共有5个等级,按由低到高分别为R1(3级)、R2(2级)、R3(1级)、R4(特3级)、R5(特2级)。

从2014年6月1日开始,各等级茶叶样品每50g分别采用锡纸袋加干燥剂独立密封包装,放入温度为5℃的冰柜中保存。

1.2 仪器设备

仪器采用PEN3型便携式电子鼻,该电子鼻由德国Airsense公司生产,系统主要组成部分包括:传感器阵列、数据采集软件、辅助功能单元(采样/清洗通道、吸附/解附单元)。传感器阵列包含10个金属氧化物传感器,各传感器特性如表1所示。传感器输出为 G/G_0 ,其中 G 为传感器接触挥发性气体的电导率, G_0 为传感器在经过活性炭过滤之后的纯净空气中的电导率。

表1 各传感器特性

Table 1 Properties of the sensors

传感器	响应特性	检测范围 (mL/m^3)
W1C(S1)	对芳香成分灵敏	10
W5S(S2)	灵敏度大,对氮氧化物很灵敏	1
W3S(S3)	对氨水、芳香成分灵敏	10
W6S(S4)	对氢气有选择性	100
W5C(S5)	对烷烃、芳香成分灵敏	1
W1S(S6)	对甲烷灵敏	100
W1W(S7)	对硫化成分灵敏,对烃和硫的有机成分较灵敏	1
W2S(S8)	对乙醇灵敏	100
W2W(S9)	对芳香成分、有机硫化物灵敏	1
W3S(S10)	对烷烃灵敏	100

1.3 试验方法

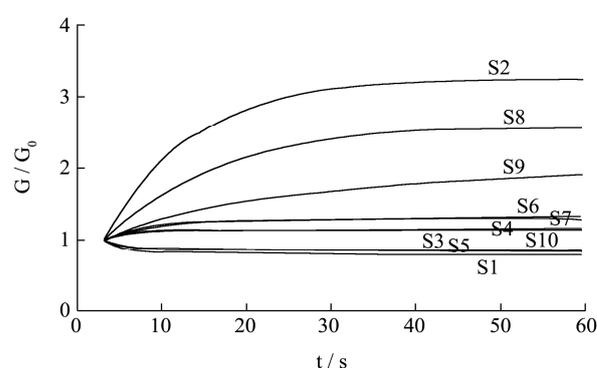


图1 电子鼻典型响应曲线

Fig.1 Typical response curve of the electronic nose

从2014年6月1日至2015年3月30日,对不同等级茶叶样品每隔60d做一次检测,连续检测300d,即0d、60d、120d、180d、240d、300d。每次试验时,室温保持在26℃,从冰柜中取出5种不同

等级的干茶叶并分别准备 45 个重复样品, 每个样品 5 g, 放入 250 mL 烧杯中双层薄膜密封, 静置 45 min。电子鼻每隔 1 s 采样一次, 连续采样 60 s, 清洗时间 50 s。

1.4 数据分析方法及建模方法

茶叶香气成分构成复杂, 各个传感器对于同一香气都有响应, 存在信息交叉。而为了尽可能全面的反映原始数据包含的信息, 选择的原始特征变量较多。因此, 对原始特征向量采用 PCA 进行降维处理, 消除冗余信息^[12], 提取出主特征向量。

神经网络可以任意精度逼近非线性函数且泛化能力强, 是模式识别中经常采用的方法。而 BPNN 是神经网络中最常用的一种, 具有结构简单、设计方便等优点。因此, 本研究采用 BPNN 建立茶叶贮藏时间的预测模型。

2 结果与讨论

2.1 原始特征向量的选择

电子鼻典型响应曲线如图 1 所示。从响应曲线可

表 2 不同贮藏时间 R2 茶叶部分变量相关系数矩阵

Table 2 Correlation coefficient matrix of partial variables of R2 with different storage times

变量	x_1	x_2	x_{11}	x_{12}	x_{21}	x_{22}	x_{31}	x_{32}
x_1	1.000							
x_2	0.163	1.000						
x_{11}	0.494	0.390	1.000					
x_{12}	0.476	0.322	-0.208	1.000				
x_{21}	0.776	-0.398	0.759	0.186	1.000			
x_{22}	0.036	0.679	0.347	0.702	0.233	1.000		
x_{31}	0.426	0.192	0.529	-0.130	0.673	0.100	1.000	
x_{32}	0.268	0.567	0.251	0.460	0.281	0.327	0.805	1.000

2.3 PCA 分析

为了克服原始数据量级的大小对主成分分析的影响, 在主成分分析前对原始特征向量进行如下标准化处理:

$$x_{ip}^* = \frac{x_{ip} - \bar{x}_p}{\sqrt{\text{Var}(x_p)}}$$

式中, \bar{x}_p 为第 p 个变量的均值; $\sqrt{\text{Var}(x_p)}$ 为第 p 个变量的标准差。

通过主成分分析, 发现各等级不同贮藏时间茶叶原始特征向量前 5 个主成分方差都大于 1, 且累计方差贡献率均超过 90%。因此, 选择前 5 个主成分作为

以看出各传感器在 60 s 以后响应趋于稳态。根据各传感器响应曲线的变化特点, 选择各传感器最大响应值、第 60 s 时的响应值 (即稳态值)、60 s 内响应的均值、60 s 内响应曲线积分值, 作为原始特征变量。则由特征变量构成 40 维的特征向量:

$$X=(x_1, x_2, \dots, x_{39}, x_{40})$$

其中, $x_1 \sim x_{10}$ 分别为 10 个传感器最大响应值; $x_{11} \sim x_{20}$ 分别为 10 个传感器第 60 s 时响应值; $x_{21} \sim x_{30}$ 分别为 10 个传感器响应 60 s 内的平均值; $x_{31} \sim x_{40}$ 分别为 10 个传感器响应 60 s 内的积分值。

2.2 相关性分析

对各等级不同贮藏时间茶叶样本原始特征向量进行相关性分析, 其中相关系数采用 Pearson 相关, 显著性检验采用 Two-tailed 检验。根据相关性分析的结果, 部分变量之间的相关系数值较大, 存在明显的相关性。如果以这些变量作为神经网络的输入, 将会降低模型的预测精度。因此, 有必要通过主成分分析消除原始特征向量中包含的重叠和冗余信息。不同贮藏时间 R2 茶叶部分变量相关系数矩阵如表 2 所示。

主特征变量可以概况原始变量的主要信息, 原始特征向量由 40 维降为 5 维。5 个等级 6 个不同贮藏时间前 5 个主成分方差及贡献率如表 3、表 7 所示。

表 3 R1 茶叶不同贮藏时间前 5 个主成分方差及贡献率

Table 3 Variance and contribution rate of the first five principal components of R1 with different storage times

主成分	方差	贡献率/%	累计贡献率/%
F_1	45.56	48.55	48.55
F_2	22.47	23.94	72.49
F_3	9.72	10.36	82.85
F_4	8.28	8.82	91.67
F_5	1.69	1.80	93.47

表 4 R2 茶叶不同贮藏时间前 5 个主成分方差及贡献率

Table 4 Variance and contribution rate of the first five principal components of R2 with different storage times

主成分	方差	贡献率/%	累计贡献率/%
F_1	47.16	46.34	46.34
F_2	32.47	31.90	78.24
F_3	6.52	6.40	84.64
F_4	5.31	5.22	89.86
F_5	2.69	2.64	92.50

表 5 R3 茶叶不同贮藏时间前 5 个主成分方差及贡献率

Table 5 Variance and contribution rate of the first five principal components of R3 with different storage times

主成分	方差	贡献率/%	累计贡献率/%
F_1	45.89	44.62	44.62
F_2	37.10	36.07	80.69
F_3	6.02	5.85	86.54
F_4	5.11	4.97	91.51
F_5	3.01	2.93	94.44

表 6 R4 茶叶不同贮藏时间前 5 个主成分方差及贡献率

Table 6 Variance and contribution rate of the first five principal components of R4 with different storage times

主成分	方差	贡献率/%	累计贡献率/%
F_1	50.76	53.17	53.17
F_2	24.32	25.47	78.64
F_3	7.35	7.70	86.34
F_4	4.13	4.33	90.67
F_5	1.01	1.06	91.73

表 7 R5 茶叶不同贮藏时间前 5 个主成分方差及贡献率

Table 7 Variance and contribution rate of the first five principal components of R5 with different storage times

主成分	方差	贡献率/%	累计贡献率/%
F_1	49.24	49.28	49.28
F_2	27.46	27.48	76.76
F_3	11.31	11.32	88.08
F_4	3.09	3.09	91.17
F_5	2.31	2.31	93.48

2.4 PCA-BPNN 预测模型设计

BPNN 由输入层、中间层和输出层组成, 网络的学习过程中通过误差反向传播不断修正连接权和阈值, 直到误差达到设定精度或学习次数达到最大设定次数学习结束。设计中采用单中间层的 3 层网络结构。输入层节点数等于输入变量 (即主成分) 个数为 5 个, 网络输出为茶叶贮藏时间预测值, 输出层节点数为 1 个。中间层节点数选择没有统一的有效方法, 设计过

程中依靠经验法在一定的单元数范围内经过多次试验进行确定。中间层激励函数选择 Tansig 函数, 输出层激励函数选择 Purelin 函数, 学习率为 0.1, 动量因子为 0.5, 误差精度设定为 0.01。将各等级每个不同贮藏时间 45 个茶叶样品分成两部分, 其中 30 个样本作为学习样本, 15 个样本作为预测样本。PCA-BPNN 不同中间层节点数学习结果如表 8 所示。可以看出当中间层为 10 个节点时平均学习次数最少为 2223 次, 性能较好。因此, PCA-BPNN 结构设计为 5-10-1。同时, 为了比较网络性能还设计了以原始特征变量为输入的 BPNN (以下称为 BPNN), 设计方法和参数设置与前面相同, 经过反复试验确定最佳网络结构为 40-18-1, 平均学习次数为 3431 次。R3 等级茶叶两种网络的学习曲线如图 2 所示。可以看出 PCA-BPNN 相比 BPNN 学习效率明显提高。

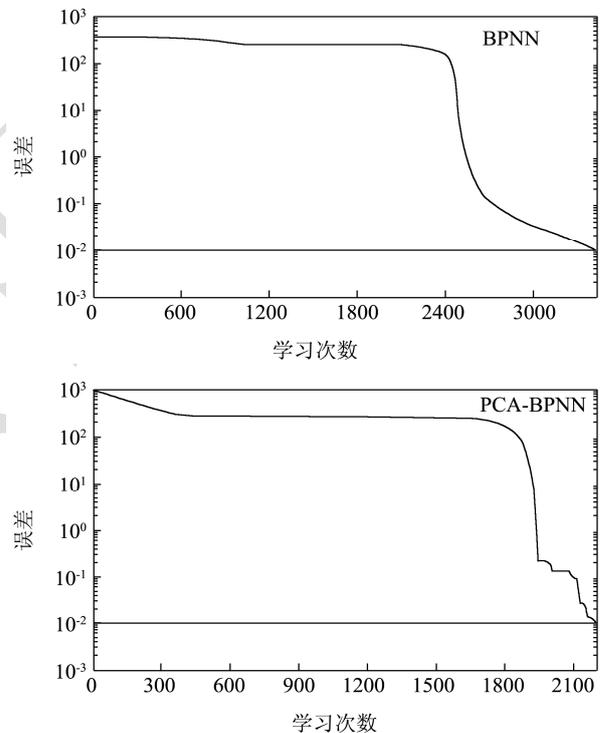


图 2 R3 等级茶叶 2 种网络的学习曲线

Fig.2 Learning curves of two networks for R3

表 8 PCA-BPNN 不同中间层节点数学习结果

Table 8 Learning results of PCA-BPNN with different node numbers in the middle layer

隐层单元数	学习次数	误差
4	4564	0.01
6	4132	0.01
7	3685	0.01
8	2345	0.01
10	2223	0.01

转下页

接上页

12	2851	0.01
15	4107	0.01
20	3981	0.01

2.5 PCA-BPNN 模型预测结果

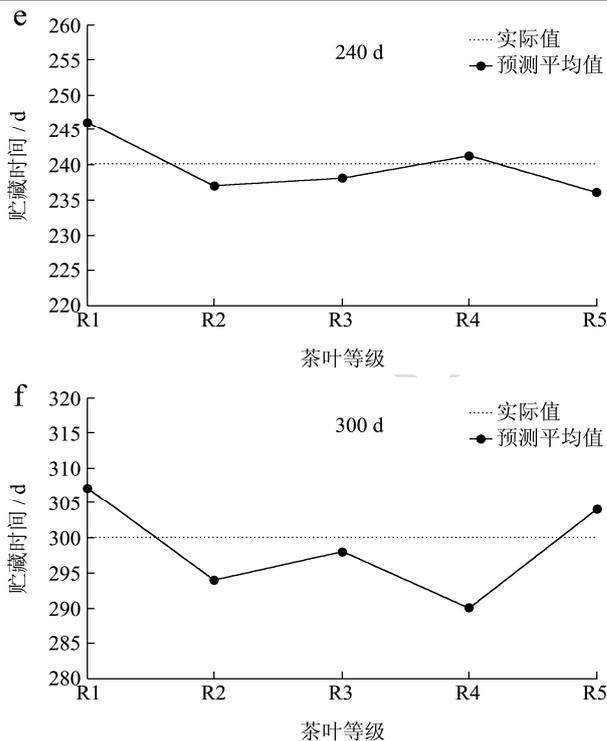
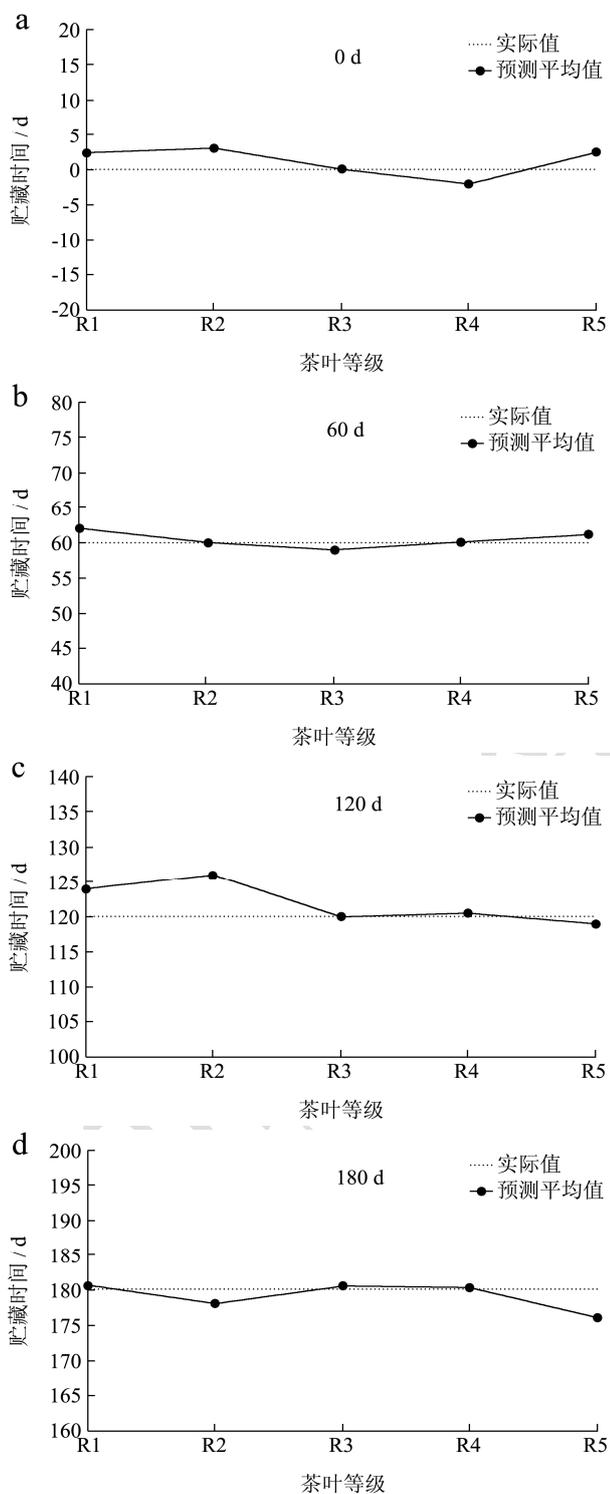


图3 PCA-BPNN 对各等级茶叶不同贮藏时间预测平均值
Fig.3 Predicted averages of each tea sample grade with different storage times using PCA-BPNN

利用 PCA-BPNN 和 BPNN 对 5 个等级每个不同贮藏时间的 15 个测试样本进行预测。2 网络对各等级茶叶贮藏时间预测结果分析如表 10 所示。从分析结果看, PCA-BPNN 最大预测误差为 21 d, BPNN 为 35 d; PCA-BPNN 预测误差超过 10 d 的样本数及比例都小于 BPNN。说明 PCA-BPNN 的预测精度优于 BPNN。PCA-BPNN 对各等级茶叶不同贮藏时间预测平均值如图 3a~f 所示。

表 10 2 网络对各等级茶叶贮藏时间预测结果分析

Table 10 Predicted storage time of each tea sample grade using both networks

网络	贮藏天数	最大预测误差天数	预测误差超过 10 d 样本	
			个数	比率/%
PCA-BPNN	0	7	5	6.67
	60	10	4	5.33
	120	16	7	9.33
	180	19	8	10.67
	240	21	8	10.67
	300	14	6	8.00
BPNN	0	12	10	13.33
	60	16	17	22.67
	120	22	20	26.67

转下页

接上页

180	31	21	28.00
240	35	19	25.33
300	33	14	18.67

3 结论

本文利用电子鼻对 5 个等级 6 个不同贮藏时间下的黄山毛峰茶进行了检测。通过 PCA 提取出不同贮藏时间各等级茶叶前 5 个主成分,以 5 个主成分作为 BPNN 的输入建立了茶叶贮藏时间预测模型。从预测结果分析,PCA-BPNN 对于 5 个等级 6 个不同贮藏时间的茶叶样本最大预测误差为 21 d。对于贮藏 0 d 的茶叶,最大预测误差为 7 d,5 个(6.67%)样本预测误差超过 10 d;对于贮藏 60 d 的茶叶,最大预测误差为 10 d,4 个(5.33%)样本预测误差超过 10 d;对于贮藏 120 d 的茶叶,最大预测误差为 16 d,7 个(9.33%)样本预测误差超过 10 d;对于贮藏 180 d 的茶叶,最大预测误差为 19 d,8 个(10.67%)样本预测误差超过 10 d;对于贮藏 240 d 的茶叶,最大预测误差为 21 d,8 个(10.67%)样本预测误差超过 10 d;对于贮藏 300 d 的茶叶,最大预测误差为 14 d,6 个(8.00%)样本预测误差超过 10 d。与以原始特征向量作为输入量建立的 BPNN 的预测模型相比,PCA-BPNN 建模效率和预测精度更高,性能更好。因此,PCA 与 BPNN 结合的方法可用于黄山毛峰茶不同贮藏时间的检测更有效,对于其他种类茶叶品质检测具有一定的借鉴意义。

参考文献

- [1] 赵菁.绿茶特征香气成分及与品质的关系研究[D].杭州:浙江大学,2002
ZHAO Jing. Study on the relationship between the quality and characteristics of green tea aroma components [D]. Hangzhou: Zhejiang University, 2002
- [2] 杨俊.电子鼻识别中的预处理与模糊神经网络算法[D].大连:大连理工大学,2007
YANG Jun. Arithmetic of signal preprocessing and fuzzy neural network in recognition of electronic nose [D]. Dalian: Dalian University of Technology, 2007
- [3] 尹芳缘,曾小燕,徐薇薇,等.基于电子鼻的芒果储存时间预测方法研究[J].传感技术学报,2012,25(9):1199-1203
YIN Fang-yuan, ZENG Xiao-yan, XU Wei-wei, et al. Study of mango storage time predicting method utilizing electronic nose [J]. Chinese Journal of Sensors and Actuators, 2012, 25(9): 1199-1203
- [4] Hui Guo-hua, Wu Yu-ling, Ye Dan-dan, et al. Study of peach freshness predictive method based on electronic nose [J]. Food Control, 2012, 28: 25-32
- [5] Liu Ming, Han Xiao-min, Tu Kang, et al. Application of electronic nose in Chinese spirits quality control and flavor assessment [J]. Food Control, 2012, 26: 564-570
- [6] Zhou Qiong, Zhang Shun-ping, Li Yu-xiao, et al. A Chinese liquor classification method based on liquid evaporation with one unmodulated metal oxide gas sensor [J]. Sensors and Actuators B: Chemical, 2011, 60(1): 483-489
- [7] 洪雪珍,韦真博,海铮,等.基于电子鼻和神经网络的牛肉新鲜度的检测[J].现代食品科技,2014,30(4):279-285
HONG Xue-zhen, WEI Zhen-bo, HAI Zheng, et al. Application of electronic nose and neural network in beef freshness detection [J]. Modern Food Science and Technology, 2014, 30(4): 279-285
- [8] Balasubramanian S, Panigrahi S, Logue C M, et al. Neural networks-integrated metal oxide-based artificial olfactory system for meat spoilage identification [J]. Journal of Food Engineering, 2009, 91(1): 91-98
- [9] 于慧春,王俊.电子鼻技术在茶叶品质检测中的应用研究[J].传感技术学报,2008,21(5):748-752
YU Hui-chun, WANG Jun. Tea quality detection basic on electronic nose [J]. Chinese Journal of Sensors and Actuators, 2008, 21(5): 748-752
- [10] 陈哲,赵杰文.基于电子鼻技术的碧螺春茶叶品质等级检测研究[J].农机化研究,2012,34(11):133-137
CHEN Zhe, ZHAO Jie-wen. Detection of biluochun green tea quality grade by using an electronic nose [J]. Journal of Agricultural Mechanization Research, 2012, 34(11): 133-137
- [11] TANG Wan-mei. The Study of the optimal structure of bp neural network [J]. System Engineering Theory and Practice, 2005, 25(10): 95-100
- [12] Lei Zhang, Rastislav L, Wu Xiao-lin. PCA-based spatially adaptive denoising of CFA images for single-sensor digitalcameras [J]. IEEE Transaction on Image Processing, 2009, 18(4): 797-812