

# 基于传感器阵列多特征优化融合的鱼粉品质检测

李培<sup>1</sup> 谭鹤群<sup>1,2</sup> 张伟健<sup>1</sup> 皇甫季璇<sup>1</sup> 牛智有<sup>1,2</sup>

1. 华中农业大学工学院, 武汉 430070; 2. 农业农村部长江中下游农业装备重点实验室, 武汉 430070

**摘要** 为了提高鱼粉品质检测装置的鉴别能力, 利用研制的鱼粉品质检测装置, 提取鱼粉样本的响应特征信息(10×6个)构成原始特征矩阵, 以多层感知器神经网络的鉴别正确率为评价指标, 对其传感器阵列进行多特征数据融合优化。首先, 通过不同的归一化处理, 得到了最佳的归一化处理方法; 其次, 通过因子载荷分析结果计算获得1770个特征距离值, 按从小到大的顺序对1770个距离进行排序, 并依据特征值距离原点的欧式距离, 剔除欧氏距离较小的19个特征值, 获得最高的鉴别正确率; 最后, 对经过载荷分析优化后的原始特征值进行相关性分析, 按相关系数绝对值累加和大小进行排序, 当剔除掉相关系数绝对值累加和大于37.2时的8个特征值时, 此时鉴别正确率为98.3%, 特征子集也更紧凑。研究表明: 特征优化前后的传感器信号的表征特征发生了明显的变化, 33个特征值被用来表征鱼粉样本的传感器特征信号。同时, 采用马氏距离解释了MLP神经网络鉴别结果的可信性, 进一步说明了特征优化方法的合理性。

**关键词** 鱼粉; 嗅觉传感器; 传感器阵列; 多层感知器神经网络; 无损检测; 归一化; 载荷分析

**中图分类号** S24 **文献标识码** A **文章编号** 1000-2421(2019)06-0163-08

随着我国水产养殖规模的不断增大, 水产饲料需求量快速增长, 对鱼粉的需求量也呈持续高增长状态。鱼粉(fish meal)蛋白质含量丰富, 并含有丰富的维生素和未知生长因子, 是饲料工业主要的动物源性饲料原料, 因此, 其品质直接影响饲料产品的质量<sup>[1]</sup>。鱼粉在储藏过程中因天气变化、温度、湿度、霉菌及其他微生物的影响, 会导致其品质和营养成分发生不同程度的变化, 从而影响动物的消化与免疫能力<sup>[2-3]</sup>。因此, 鱼粉品质的检测是确保饲料产品质量的重要技术环节和手段。

电子鼻技术是近年来发展起来的一种新的检测技术, 与常规理化分析方法(如化学检测、感官评定、光谱分析等)相比<sup>[4]</sup>, 其具有无需复杂的前处理、不使用有污染的化学试剂以及简单的操作过程等优点。目前, 该项技术应用广泛, 如赵梦醒等<sup>[5]</sup>将电子鼻应用在鲈鱼新鲜度检测当中, 区分度较好; Kriengkri等<sup>[6]</sup>将电子鼻技术应用于鸡肉菌落检测, 准确率较高。因此, 电子鼻作为一种样本品质的检测工具, 具有非常大的潜力。在电子鼻检测系统中起主要作用的是气体传感器阵列, 传感器阵列特征产生的冗余信息和不相关信息对鉴别结果也极为不

利。因此, 采用电子鼻进行鱼粉样本品质检测鉴别时, 传感器特征的提取与表征特征的优化尤为重要。在已有的报道中, 针对特征提取问题, 常用的方法有很多, 但这些方法因研究对象的差异而存在应用效能的不同。Shi等<sup>[7]</sup>通过遗传算法对用于检测西湖龙井品质的传感器特征值进行优化, 优化后的检测准确率提高了3%以上。Yin等<sup>[8]</sup>通过威尔克斯统计量对用于3种食用醋种类进行分类的传感器阵列进行优化, 优化效果较好。Sajad等<sup>[9]</sup>通过聚类分析对用于藏红花分类的特征值进行了优化。徐克明等<sup>[10]</sup>基于非搜索性特征优化方法, 对用于山核桃陈化时间的电子鼻传感器阵列进行优化, 优化后的山核桃组内聚集度增大, 组间更易区分。但上述方法的实用性皆因不同的鉴别目标而存在应用效能上的差异。在已有的利用电子鼻技术检测鱼粉品质的研究中, 缺少对传感器阵列的优化, 且市场上商用电子鼻的传感器阵列没有针对性的筛选优化, 多余的传感器会导致电子鼻检测数据冗余, 增加模型的复杂度, 造成时间的浪费和模型的不稳定, 进而影响检测结果<sup>[11]</sup>。因此, 必须采用合理的适用于样本的特征优化方法, 才更有可能剔除不相关的特征或者传感

收稿日期: 2019-01-24

基金项目: 中央高校基本科研业务费专项(2662018PY081)

李培, 博士研究生, 研究方向: 农产品加工技术与装备. E-mail: huanonglipei8@163.com

通信作者: 牛智有, 博士, 教授, 研究方向: 农产品加工技术与装备. E-mail: nzhy@mail.hzau.edu.cn

器,减少冗余信息,获得对分类最有效的特征。

传感器响应曲线的积分值、平均微分值、稳定值常被用来表征样本信息<sup>[12]</sup>,但 3 种特征并不足以表征样本所有的动态信息和稳态信息,还需找到其他的特征参量,因此,能量值、最大梯度值、方差值作为补充,共同表征不同储藏时间的鱼粉样本。本研究以多层感知器神经网络(multilayer perceptron neural network)的鉴别正确率为指标,以传感器阵列特征为研究对象,通过归一化方法优化、基于因子之间距离的因子载荷分析/loading)优化以及相关性分析优化,对大量冗余信息进行剔除。为验证该选择策略的有效性,采用马氏距离来解释 MLP 神经网络鉴别结果的可信性,得到最终优化后的传感器阵列,为传感器阵列多特征融合的鱼粉品质检测提供一种新的优化方法。

## 1 材料与方法

### 1.1 试验材料

取新鲜的鱼粉储藏于高温高湿环境中,使鱼粉在储藏过程中逐渐变质,总共获得了 6 个不同储藏

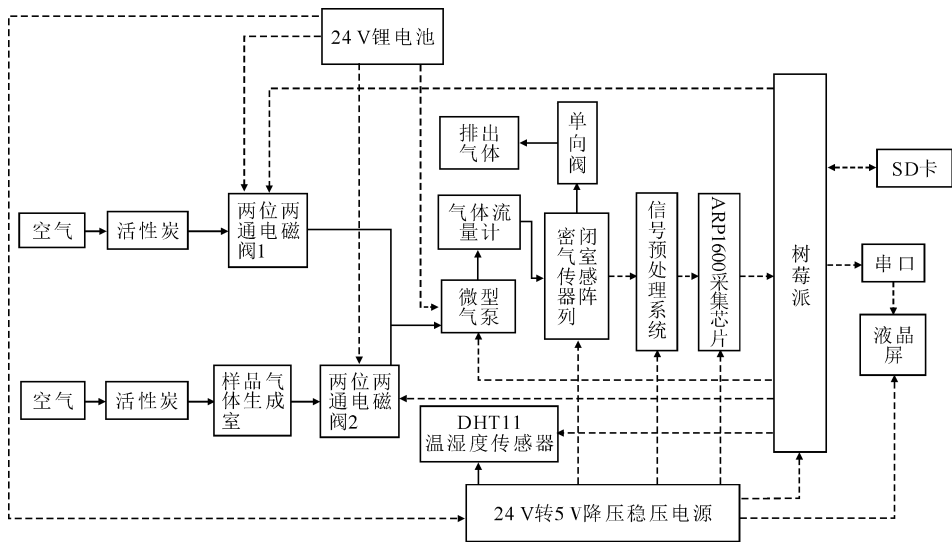
时间的样本,每种样本平均分成 30 份,每份质量为 80 g,总共进行 180 次试验;其中,122 个样本作为训练集,用于建立模型;58 个样本构成测试集,用于验证模型。

### 1.2 电子鼻检测系统

试验采用的系统为自行研制的便携式鱼粉品质检测系统。该系统组成如图 1 所示,主要由气体采集与传输模块、以树莓派为核心的控制处理存储模块、ARPI600 数据采集模块、传感器阵列模块组成。传感器阵列模块为该装置的核心部件,主要由 10 个气体传感器组成,即传感器 1(TGS822)、传感器 2(TGS2602)、传感器 3(TGS813)、传感器 4(TGS2620)、传感器 5(MQ136)、传感器 6(TGS2600)、传感器 7(MQ139)、传感器 8(TGS2610)、传感器 9(MQ137)、传感器 10(TGS2611)。

### 1.3 试验方法

首先对该检测装置进行预热,以期消除仪器不稳定所造成的影响,再用经活性炭过滤之后的纯净空气对装置进行清洗,清洗时间为 77 s,最后将鱼粉样本置于 250 mL 的高硼硅采样瓶中,由微型气泵



实线代表气路,虚线代表电路。The solid line represented gas paths, and the dotted line represented circuits.

图 1 检测装置的系统组成框图

Fig.1 Block diagram of detection device system

将样本产生的顶空气体抽至该检测装置的气体采样室中,与位于采样室中的气体传感器发生化学反应,进而引起传感器的阻值发生变化,该装置的检测时间为 39 s,数据采样间隔为 1 s。每次采样完之后都需要对装置进行清洗复原,以便进行下个样本的测试。各个传感器对腐败样本的响应曲线如图 2 所

示。从图 2 可以看出,每个传感器对鱼粉样本都有响应,且响应趋势大致相同,都是随着传感器表面富集的样本气体的不断增大,电导比值呈现先增大后趋于平缓,达到一个相对稳定的状态。所以在后期特征值优化中,对每一个传感器的特征值都将作为初始特征值来优化。

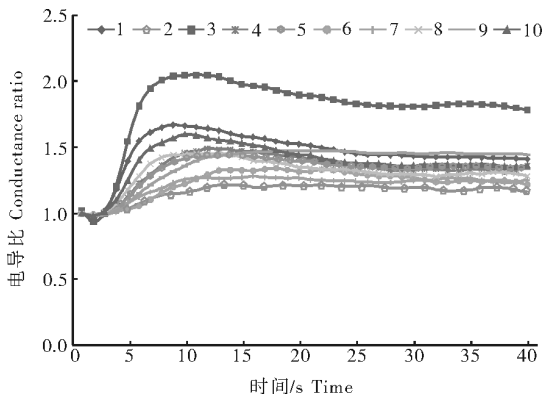


图 2 腐败鱼粉的传感器响应曲线

Fig.2 Sensor response curve of corrupt fish meal

1.4 数据处理

采用 Savitzky-Golay 进行五点滤波平滑处理<sup>[13]</sup>消除噪声信号的影响,提取积分值(integral value, INV)、能量值(wavelet energy value, WEV)、最大梯度值(maximum gradient value, MGV)、平均微分值(average differential value, ADV)、相对稳定均值(relation steady-state response average value, RSAV)、方差值(variance value, VARV) 6 种特征作为鱼粉样本电子鼻信号的特征参数<sup>[14]</sup>,各特征参数计算如式(1)~(6)所示,共提取 10×6 个特征参数,构成 60 维特征矩阵。由于 6 种特征值的量纲不同,需对各特征值进行数据归一化处理。各特征值编号与传感器编号的对应关系如表 1 所示。以多层感知器神经网络鉴别正确率<sup>[15]</sup>为标准,其中,多层感知器神经网络的隐藏层激活函数为双曲正切函数,输出层激活函数为 Softmax 函数,通过因子载荷分析、相关性分析对原始特征矩阵进行优化分析,得到优化后的传感器特征矩阵。

1) 积分值(INV)。积分值是传感器响应信号曲线与其基线间的面积,反映了该传感器对待测对象挥发性成分的总体响应结果。其计算公式如下:

$$I = \sum_{i=1}^N X_i \Delta t \tag{1}$$

2) 能量值(WEV)。采用 db3 小波函数对传感器原始数据进行 4 层小波分解,每个传感器得到 7 个近似系数作为特征,将每个系数的平方和作为每个传感器的能量值。

$$W = \sum_{i=1}^m a_{i}^2 \tag{2}$$

3) 最大梯度值(MGV)。气敏传感器响应曲线存在 1 个最大响应值,因此可以用最大值与初始值之间的线段斜率来表示初始阶段信号变化快慢。

$$M = \frac{x_{i_{max}} - x_0}{i} \tag{3}$$

4) 平均微分值(ADV)。平均微分值是可以全面反映传感器动态响应过程整体信息的一种方法,直接体现了气敏传感器对气体响应的主流信息。

$$A = \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{x_{i+1} - x_i}{\Delta t} \tag{4}$$

5) 相对稳定均值(RSAV)。气敏传感器响应曲线存在 1 个相对稳态区间,因此可用此区间的平均值来表征稳态特征,其计算公式如下:

$$R = \frac{\sum_{i=t_0}^N x_i}{N - t_0} \tag{5}$$

6) 方差值(VARV)。方差反应了数据的离散程度,利用响应信号的方差来表示其信号强度的特征值。

$$V = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N} \tag{6}$$

式(1)~(6)中: $x_i$ 为传感器对 1 个样本第  $i$  秒的响应值; $\bar{x}$ 为传感器对 1 个样本响应信号均值; $\Delta t$ 为相邻 2 个采样点的时间间隔,取 1 s; $t_0$ 为曲线即将稳定时所对应的时间,s; $a_{i}$ 为信号四尺度分解后逼近系数集中第  $i$  个分解系数; $m$ 为逼近系数集中的系数总数; $N$ 为传感器对 1 个样本的采集时间; $x_0$ 为传感器对 1 个样本的初始响应值。

表 1 各特征编号与传感器编号的对应关系

Table 1 Corresponding relation of sensor serial and each feature number

特征参数 Feature parameter	特征值编号 Feature number	传感器编号 Sensor number
积分值 INV	$I_1, I_2, I_3, I_4, I_5, I_6 \cdots I_{10}$	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
能量值 WEV	$W_1, W_2, W_3, W_4, W_5, W_6 \cdots W_{10}$	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
最大梯度值 MGV	$M_1, M_2, M_3, M_4, M_5, M_6 \cdots M_{10}$	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
平均微分值 ADV	$A_1, A_2, A_3, A_4, A_5, A_6 \cdots A_{10}$	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
相对稳定均值 RSAV	$R_1, R_2, R_3, R_4, R_5, R_6 \cdots R_{10}$	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
方差值 VARV	$V_1, V_2, V_3, V_4, V_5, V_6 \cdots V_{10}$	1, 2, 3, 4, 5, 6, 7, 8, 9, 10

## 2 结果与分析

### 2.1 原始特征矩阵归一化处理

本研究选取了 6 种归一化方法进行分类比较,各归一化方法公式如式(7)~(12)所示,并根据多层感知器神经网络的鉴别正确率选出分类精度较高的归一化方法作为本研究的归一化方法。

$$x'_i = \frac{x_i}{x_{\max} - x_{\min}} \quad (7)$$

$$x'_i = \ln x_i \quad (8)$$

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (9)$$

$$x'_i = x_i / \sum x_i \quad (10)$$

$$x'_i = \lg x_i \quad (11)$$

$$x'_i = \frac{2(x_i - x_{\min})}{x_{\max} - x_{\min}} - 1 \quad (12)$$

式(7)~(12)中,  $x'_i$  为归一化后的特征值;  $x_i$

为原始数据特征值;  $x_{\max}$  为原始特征最大值;  $x_{\min}$  为原始特征最小值。

各方法的归一化范围以及区间长度和分类精度如表 2 所示。由表 2 可知,未进行归一化的原始数据的分类精度只有 33.3%,并不能达到分类标准,而经过归一化处理的数据进行分类得到的鉴别正确率均比未经过归一化处理的正确率高,不同的归一化方法又有不同的分类精度。从表 2 可以看出,第 3 种方法的分类精度最高,为 96.7%,而此分类方法的区间长度为 1。区间长度最小的第 4 种归一化方法的鉴别正确率为 53.3%,区间长度最长的第 2 种归一化方法的鉴别正确率仅为 56.7%,所以,区间长度直接影响着数据集的分布稀疏程度,且归一化区间长度并不是越大越好,范围过大,则会丧失归一化的优势。因此,区间长度为 1 的第 3 种归一化方法对于不同品质的鱼粉数据集来说有更好的分类精度。

表 2 各归一化方法的分类精度

Table 2 Comparison of classification accuracy of each normalization method

归一化方法 Normalization method	归一化范围 Normalized range	区间长度 Interval length	鉴别正确率/% Correct discrimination rate
原始特征 Original feature	/	/	33.3
1	[0,6]	6	94.4
2	[-12,6]	18	56.7
3	[0,1]	1	96.7
4	[0,0.2]	0.2	53.3
5	[-5,3]	8	58.3
6	[-1,1]	2	95.0

### 2.2 载荷分析

载荷分析结果如图 3 所示,横坐标为第一主成分,纵坐标为第二主成分,从各个特征变量的坐标分布来看,每个特征值离坐标原点的距离都较远,说明这些特征值对样本检测都有贡献。但从图 3 可知,大量特征变量的因子载荷分布较接近,说明它们对

样本的检测具有一致的贡献作用,而欧式距离经常用于表征样本点间的相似程度,距离越近越相似。因此,可以考虑在重叠比较大的特征变量中,依据距离原点的欧式距离,剔除欧氏距离较小的特征值,以减小冗余信息。

从两两之间总共 1 770 个距离中适当地剔除较小的距离值,剔除的距离值的大小不同,导致剔除的原始特征变量的数目不同,进而引起样本的鉴别结果也发生变化。剔除的特征值太多会导致表征样本的特征减小,从而使得分类结果下降,故只考虑剔除 6 个距离段的特征值。剔除的距离与原始特征变量数目的对应关系如表 3 所示。

图 4 为剔除不同的原始特征数目所对应的多层感知器神经网络的鉴别正确率。从图 4 可知,当剔除的原始特征变量数目为 19 个时,即剔除距离小于 0.008 4,此时鉴别正确率最高,为 97.8%。但此时选择的特征数目为 41 个,相对较多。当剔除的原始特征变量为 32 个时,此时相对于最初的变量数目,

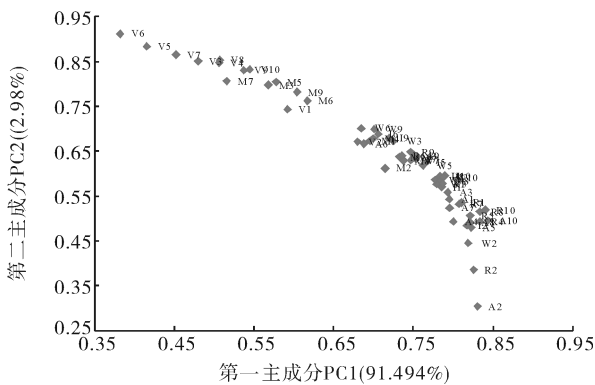


图 3 因子载荷分布图

Fig.3 Factor loadings distribution diagram

已经剔除了 53.3% 的变量, 此时的鉴别正确率已低于 90%, 所以需继续采用其他特征优化方法继续探索得出最佳的分类准确率和特征数目。

表 3 剔除的距离与原始特征变量数目的对应关系

Table 3 Relationship between the removed distance and the number of original feature variables

剔除的距离 Removed distance	原始特征变量数目 Number of original feature variables
<0.004 6	5
<0.006 5	8
<0.008 4	19
<0.010 3	23
<0.012 2	28
<0.014 1	32

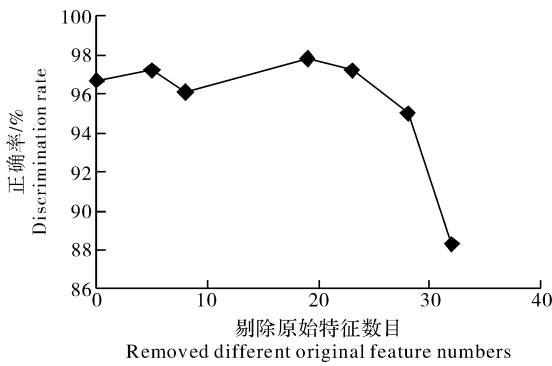


图 4 剔除不同的原始特征数目所对应的鉴别正确率的变化

Fig.4 Relationship between correct discrimination rates and removed different original feature numbers

### 2.3 相关性分析

传感器之间的相关程度用 Pearson 相关系数来表示:

$$R_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (13)$$

式(13)中,  $x_i$  为特征值  $x$  对第  $i$  个样本的相对变化值;  $y_i$  为特征值  $y$  对第  $i$  个样本的相对变化值;  $\bar{x}$  为特征值  $x$  对所有样本的均值;  $\bar{y}$  为特征值  $y$  对所有样本的均值。

由于无法从任意 2 个特征值的相关度推知整体整列的相关程度, 故引用某个特征值与其他特征值的相关系数绝对值累加和  $r_i$  来判断,  $r_i$  计算如式(14)所示。对经过载荷分析优化后的 41 个特征值进行相关性分析, 相关系数绝对值累加和如表 4 所示。

$$r_i = \sum_{j=1}^{41} |R_{xy}| \quad (14)$$

$r_i$  值越小, 则该特征值与其他类特征值的相关度越低, 但是具体选择多大的相关系数绝对值累加和作为节点, 则需要通过试验来确定。本研究以多层感知神经网络对样本的鉴别正确率为标准, 对剔除的相关系数绝对值累加和节点进行试验, 最终得出当剔除掉相关系数绝对值累加和大于 37.2 时, 此时的鉴别正确率最高, 为 98.30%, 相比于因子载荷优化分析后, 鉴别正确率提高了, 剔除的特征值分别为  $I_3$ 、 $I_9$ 、 $W_3$ 、 $W_9$ 、 $I_5$ 、 $I_6$ 、 $I_{10}$ 、 $R_9$ 。剔除不同的原始变量数对应的鉴别正确率如图 5 所示。从图 5 可以看出, 剔除相关系数绝对值累加和大于 37.43 的特征值(剔除 4 个特征值)与剔除掉相关系数绝对值累加和大于 37.2 的特征值(剔除 8 个特征值)的鉴别

表 4 41 个特征值的相关系数绝对值累加和

Table 4 Cumulative absolute value of correlation coefficients for each sensor

特征值编号 Feature number	数值 Value	特征值编号 Feature number	数值 Value	特征值编号 Feature number	数值 Value	特征值编号 Feature number	数值 Value
$I_3$	37.433 1	$M_4$	36.878 5	$A_6$	36.287 7	$V_1$	35.536 4
$I_5$	37.402 7	$M_5$	36.851 3	$A_7$	35.623 4	$V_2$	36.398 3
$I_6$	37.364 4	$M_6$	36.814 8	$A_8$	35.469 3	$V_3$	35.353 5
$I_9$	37.478 5	$M_7$	35.185 9	$A_{10}$	36.211 3	$V_5$	34.399 9
$I_{10}$	37.267 5	$M_8$	36.653 7	$R_1$	36.240 9	$V_6$	34.147 4
$W_2$	34.429 8	$M_9$	36.982 0	$R_2$	33.093 0	$V_7$	34.934 7
$W_3$	37.493 6	$A_1$	35.988 1	$R_4$	35.792 4	$V_8$	36.155 6
$W_6$	37.111 5	$A_2$	31.105 7	$R_5$	35.876 0	$V_{10}$	36.633 6
$W_9$	37.484 8	$A_3$	36.433 5	$R_6$	36.965 6		
$M_2$	35.738 0	$A_4$	34.894 9	$R_9$	37.427 8		
$M_3$	36.364 7	$A_5$	35.223 5	$R_{10}$	36.730 3		

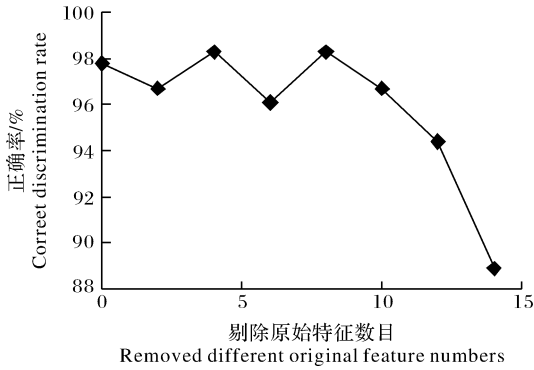


图 5 剔除不同原始变量数对应的鉴别正确率

Fig.5 Relationship between correct discrimination rates and removed different original feature numbers

表 5 特征值优化后的各传感器的型号、特征值与特征编号的对应关系  
Table 5 Relation of each sensor serial and feature number in optimized sensor array

传感器型号 Sensor	特征值编号 Feature number	特征值 Feature value
TGS822	A <sub>1</sub> , R <sub>1</sub> , V <sub>1</sub>	平均微分值、相对稳定均值、方差值 ADV, RSAV, VARV
TGS2602	W <sub>2</sub> , M <sub>2</sub> , A <sub>2</sub> , R <sub>2</sub> , V <sub>2</sub>	能量值、最大梯度值、平均微分值、相对稳定均值、方差值 WEV, MGV, ADV, RSAV, VARV
TGS813	M <sub>3</sub> , A <sub>3</sub> , V <sub>3</sub>	最大梯度值、平均微分值、方差值 MGV, ADV, VARV
TGS2620	M <sub>4</sub> , A <sub>4</sub> , R <sub>4</sub>	最大梯度值、平均微分值、相对稳定均值 MGV, ADV, RSAV
MQ136	M <sub>5</sub> , A <sub>5</sub> , R <sub>5</sub> , V <sub>5</sub>	最大梯度值、平均微分值、相对稳定均值、方差值 MGV, ADV, RSAV, VARV
TGS2600	W <sub>6</sub> , M <sub>6</sub> , A <sub>6</sub> , R <sub>6</sub> , V <sub>6</sub>	能量值、最大梯度值、平均微分值、相对稳定均值、方差值 WEV, MGV, ADV, RSAV, VARV
MQ139	M <sub>7</sub> , A <sub>7</sub> , V <sub>7</sub>	最大梯度值、平均微分值、方差值 MGV, ADV, VARV
TGS2610	M <sub>8</sub> , A <sub>8</sub> , V <sub>8</sub>	最大梯度值、平均微分值、方差值 MGV, ADV, VARV
MQ137	M <sub>9</sub>	最大梯度值 MGV
TGS2611	A <sub>10</sub> , R <sub>10</sub> , V <sub>10</sub>	平均微分值、相对稳定均值、方差值 ADV, RSAV, VARV

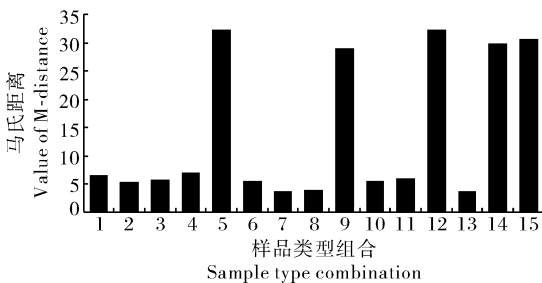
2.4 马氏距离分析验证

马氏距离是重要的类别可分性的量度指标,马氏距离越大,类别可分性越强<sup>[16-17]</sup>。马氏距离计算公式如下:

$$M_{(X,Y)} = \sqrt{(X-Y)^T \Sigma^{-1} (X-Y)} \quad (15)$$

式(15)中,  $M_{(X,Y)}$  为样本 X 与样本 Y 的马氏距离值;  $\Sigma$  为协方差矩阵,  $\Sigma > 0$ 。

从图 6 可以看出,第四类和第五类,第二类和第



1.M12; 2.M13; 3.M14; 4.M15; 5.M16; 6.M23; 7.M24; 8.M25; 9.M26; 10.M34; 11.M35; 12.M36; 13.M45; 14.M46; 15.M56.

图 6 不同类别鱼粉样本间的马氏距离值  
Fig.6 Value of M-distance between fish meal samples type combination

正确率一致,同为 98.3%,但就特征子集来说,剔除 8 个特征值的特征子集更紧凑。因此,选择剔除相关系数绝对值累加和大于 37.2 的特征值。当剔除的特征值数目过多时,使得其有效信息减少,此时的特征值已经不能表征样本信息,鉴别正确率也急速下降。经过相关性分析优化之后的特征值为 33 个。

由表 5 可知,特征值优化后的传感器的表征特征发生了很大的变化,积分值这一特征值在优化的过程中全部剔除,且没有一个特征值是完全适用于所有传感器的。因此,在鱼粉样本的鉴别分析中,传感器的特征选择是非常重要的,选择合适的特征值既有利于提高样本的鉴别能力,又能提高模型的泛化能力,提高系统的稳定性。

表 5 特征值优化后的各传感器的型号、特征值与特征编号的对应关系  
Table 5 Relation of each sensor serial and feature number in optimized sensor array

四类,第二类和第五类的马氏距离值相对于其他值来说较低,说明这些类别间的鱼粉样本较难区分。从图 6 也能看出,储藏时间等级为 6 的鱼粉样本与较低储藏时间的鱼粉样本的马氏距离比较大,区分则比较明显。

经过因子载荷优化分析和相关性分析优化之后特征值的 MLP 神经网络鉴别错分结果如表 6 所示。由表 6 可知,训练集的鉴别正确率为 100%,测试集的鉴别正确率为 94.8%。其中在验证集中,1 个第二类样本错分为第五类样本,1 个第四类样本错分为第二类样本,1 个第五类样本错分为第四类样本,其他类样本则分类正确。由此可知,类别错分情况与马氏距离分析情况完全一致,该结果解释了 MLP 神经网络鉴别正确率未达到 100% 的原因,也间接说明了特征优化方案的合理性。

3 讨论

本研究利用自行研制的鱼粉品质检测装置,对不同储藏时间等级的鱼粉样本进行检测,根据获得的响应曲线,提取 6 种特征值组成初始特征矩阵,通

表 6 MLP 神经网络的鉴别结果

Table 6 Misclassification results of MLP neural network

样本类别 Sample type	训练集 Training			验证集 Testing		
	正确分类数 Correct classified numbers	错误分类数 Misclassified numbers	错分结果 Misclassified results	正确分类数 Correct classified numbers	错误分类数 Misclassified numbers	错分结果 Misclassified results
第一类 First kind	20	0	/	10	0	/
第二类 Second kind	19	0	/	10	1	第五类 Fifth kind
第三类 Third kind	22	0	/	8	0	/
第四类 Fourth kind	21	0	/	8	1	第二类 Second kind
第五类 Fifth kind	21	0	/	8	1	第四类 Fourth kind
第六类 Sixth kind	19	0	/	11	0	/

过多层感知器神经网络方法得到的鉴别正确率为指标,并对响应特征进行归一化处理、因子载荷分析优化和因子之间的相关性分析优化。运用这一特征选择的优化方法,最终选取了 33 个原始特征值作为特征矩阵来表示不同储藏时间等级的鱼粉样本。同时,计算各储藏时间等级样本之间的马氏距离,以此对 MLP 神经网络鉴别正确率未达到 100% 的原因进行说明,更好地解释了 MLP 神经网络鉴别结果的可靠性,也说明了采用该方法进行特征选择的可行性。

由于每种特征值的量纲不同,为了消除数据属性间的差别,避免大数值的数据变化掩盖掉小数值的变化,需对数据进行归一化处理。归一化处理是一种无量纲的处理手段,可减小计算量,方便数据的处理。但是不同的归一化方法得到的分类正确率不同,究其原因,可能是由于不同的归一化方法将数据集中太大或太小的样本映射到了统一的范围内,而这个范围的区间长度直接影响着分类精度,过大或过小的区间长度可能都不能得到较高的分类精度,这个便需要试验进行确定。潘磊庆等<sup>[18]</sup>采用 Z-标准化、自动缩放、正规化、对数化等 4 种预处理方法进行归一化,得到采用 Z-标准化归一化方法能较好地地区分鸡蛋新鲜度差异,与本研究结果不同,原因是因为物料和采用的检测装置不同。

载荷分析是主成分分析中常用的一种重要方法,其目的是通过各特征值在横坐标和纵坐标的投影大小选出对分类结果贡献较大的特征值,剔除冗余的特征值,提高分类精度。相关性分析即计算各原始特征变量的相关程度,两个原始特征变量的相关系数越大,说明这 2 个特征变量所包含的样本信息的一致性就越大,则这两个特征变量可以相互取代,此时可以考虑将其中一个特征值进行剔除。但仅仅通过载荷分析或者相关性分析来决定最终特征值的选择达不到最优,必须通过多层筛选得到适合

检测鱼粉品质的传感器阵列的优化阵列。因此,本研究首先通过载荷分析选出一、二主轴上投影大的因子,即贡献大的因子,同时对冗余信息进行第一次筛选,剔除部分重合因子,再通过相关性分析对冗余信息进行第二次筛选,再次剔除重合因子,得到 33 个特征值作为最终的特征矩阵,此时的鉴别正确率为 98.3%。本研究给出的特征矩阵阵列可能不是最佳阵列,所提取的 6 类特征值也并非最适合品质的鉴别,这些问题还有待进一步的研究。

## 参 考 文 献

- [1] 王铨静.不同储存方法对鱼粉的影响及养殖效果评价[D].湛江:广东海洋大学,2017.
- [2] 孙志强,过世东.储存期间鱼粉中脂质的劣变规律[J].中国粮油学报,2015,30(3):90-93.
- [3] 王杰,梁旭方,李皎荣,等.菜粕替代鱼粉对翘嘴鲌肠道吸收和氨基酸代谢的影响[J].华中农业大学学报,2018,37(4):93-101.
- [4] 曹小华,蔡懋成,余维三,等.近红外光谱分析技术在鱼粉新鲜度检测中的应用研究[J].广东饲料,2018,27(3):42-45.
- [5] 赵梦醒,丁晓敏,曹荣,等.基于电子鼻技术的鲈鱼新鲜度评价[J].食品科学,2013,34(6):143-147.
- [6] KRIENGGKRI T, THEERAPHOP T, NOPPON L, et al. Evaluation of bacterial population on chicken meats using a briefcase electronic nose[J]. Biosystems engineering, 2016, 151: 116-125.
- [7] SHI B L, ZHAO L, ZHI R C, et al. Optimization of electronic nose sensor array by genetic algorithms in Xihu-Longjing Tea quality analysis [J]. Mathematical and computer modelling, 2013, 58: 752-758.
- [8] YIN Y, YU H C, CHU B, et al. A sensor array optimization method of electronic nose based on elimination transform of Wilks statistic for discrimination of three kinds of vinegars[J]. Journal of food engineering, 2014, 127: 43-48.
- [9] SAJAD K, SAEID M, MAHDI G V. A portable electronic nose as an expert system for aroma-based classification of saffron [J]. Chemometrics and intelligent laboratory systems, 2016,

- 156:148-156.
- [10] 徐克明, 王俊, 邓凡霏, 等. 用于山核桃陈化时间检测的电子鼻传感器阵列优化[J]. 农业工程学报, 2017, 33(3): 281-287.
- [11] 王智凝, 郑丽敏, 方雄武, 等. 电子鼻传感器阵列优化对猪肉新鲜度法的检测[J]. 肉类研究, 2015, 29(5): 27-30.
- [12] 程绍明, 王俊, 王永维, 等. 基于电子鼻技术的不同特征参数对番茄苗早疫病病害区分效果影响的研究[J]. 传感技术学报, 2014, 27(1): 1-5.
- [13] 殷勇, 郝银凤, 于慧春. 基于多特征融合的电子鼻鉴别玉米霉变程度[J]. 农业工程学报, 2016, 32(12): 254-260.
- [14] 殷勇, 吴文凯, 于慧春. 独立分量分析融合小波能量阈值的电子鼻信号去漂移方法[J]. 农业工程学报, 2014, 30(24): 325-331.
- [15] SZECOWKA P M, SZCZUREK A, LICZNERSKI B W. On reliability of neural network sensitivity analysis applied for sensor array optimization[J]. *Sensors & actuators: B. Chemical*, 2011, 157(1): 298-303
- [16] 洪雪珍, 王俊. 牛肉品质检测中电子鼻参数的优化[J]. 农业机械学报, 2013, 44(9): 125-131.
- [17] 李江勇. 基于 ARM9 的嵌入式电子鼻系统研究与实现[D]. 广州: 广东工业大学, 2012.
- [18] 潘磊庆, 刘明, 詹歌, 等. 数据预处理在电子鼻评价鸡蛋新鲜度中的应用研究[J]. 上海理工大学学报, 2010, 32(6): 584-588.

## Detecting fishmeal quality based on multi-feature optimization fusion of sensor array

LI Pei<sup>1</sup> TAN Hequn<sup>1,2</sup> ZHANG Weijian<sup>1</sup> HUANGFU Jixuan<sup>1</sup> NIU Zhiyou<sup>1,2</sup>

1. *College of Engineering, Huazhong Agricultural University, Wuhan 430070, China;*

2. *Key Laboratory of Agricultural Equipment in Mid-Lower Yangtze River, Ministry of Agriculture and Rural Affairs, Wuhan 430070, China*

**Abstract** In order to improve the identification ability of the fish meal quality detection device, the response characteristic information ( $10 \times 6$ ) of the fish meal sample is extracted to form the original feature matrix, and the multi-feature data fusion optimization of the sensor array is carried out by taking the identification accuracy of the multi-layer perceptron neural network as the evaluation index. Firstly, through different normalization processing, the best normalization processing method is obtained. Secondly, 1 770 characteristic distances are calculated by factor load analysis results, and 1 770 distances are sorted according to the order from small to large. According to the Euclidean distance between the eigenvalues and the origin, 19 eigenvalues with smaller Euclidean distance are eliminated to obtain the highest discriminant accuracy. The original eigenvalues optimized by load analysis are correlated and sorted according to the absolute sum and size of correlation coefficients. When 8 eigenvalues are removed when the absolute sum of correlation coefficients is greater than 37.2, the recognition accuracy is 98.3%, and the feature subset is more compact. The results showed that the characterization characteristics of the sensor signals changed obviously before and after feature optimization. 33 eigenvalues were used to characterize the sensor characteristic signals of fish meal samples. At the same time, the reliability of MLP neural network identification results is explained by Mahalanobis distance, which further explains the rationality of feature optimization method.

**Keywords** fish meal; olfactory sensor; sensor array; multilayer perceptron neural network; nondestructive testing; normalization; load analysis

(责任编辑: 陆文昌)