

基于气味检测的红富士苹果新鲜度识别方法研究^{*}

杨明丽 纠海峰 邓 薇

(哈尔滨商业大学轻工学院 哈尔滨 150028)

摘 要:食品安全是人民健康的基石,为实现对苹果新鲜度的快速、无损检测,提出一种应用电子鼻系统检测红富士苹果新鲜度的识别方法,该电子鼻系统是由数据采集系统和模式识别系统两部分构成,分别实现对苹果气味的浓度数据采集及数据分析。利用主成分分析、K-means 聚类、反向传播神经网络和鲸鱼算法优化的反向传播神经网络多种学习算法对采集数据进行分析。实验结果表明,基于主成分分析法降维特征的 K-means 聚类算法识别准确率为 70.67%,其值优于基于原始特征的 K-means 聚类算法识别率。利用鲸鱼算法优化的反向传播神经网络识别精度可达到 95%,可以准确识别不同新鲜度的样本,该识别方法高效、快速、便捷,可行性较强。

关键词:苹果新鲜度;电子鼻系统;气味识别;学习算法

中图分类号: TN307 **文献标识码:** A **国家标准学科分类代码:** 621.39

Research on the freshness recognition method of red Fuji apples based on odor detection

Yang Mingli Jiu Haifeng Deng Wei

(School of Light Industry, Harbin University of Commerce, Harbin 150028, China)

Abstract: Food safety is the cornerstone of people's health. To achieve rapid and non-destructive detection of apple freshness, a recognition method for detecting the freshness of Red Fuji apples using an electronic nose system is proposed. The electronic nose system consists of two parts: A data acquisition system and a pattern recognition system, which respectively collect and analyze concentration data of apple odor. Using various learning algorithms such as principal component analysis, K-means clustering, back propagation neural network, and whale algorithm optimized back propagation neural network to analyze the collected data. The experimental results show that the recognition accuracy of K-means clustering algorithm based on principal component analysis for dimensionality reduction features is 70.67%, which is better than the recognition rate of K-means clustering algorithm based on original features. The recognition accuracy of back propagation neural network optimized by whale optimization algorithm can reach 95%, which can accurately identify samples with different freshness levels. This recognition method is efficient, fast, convenient, and highly feasible.

Keywords: apple freshness; electronic nose system; odor recognition; learning algorithm

0 引 言

苹果作为人们日常生活中最常见的水果之一,在我国的产量和出口量均位居世界前列^[1],它的品质自然成为人们关注的焦点。其中红富士苹果的产量占我国苹果总产量的 72.7%^[2],满足了我国鲜食苹果消费市场周年供应的需求^[3],而且它芳香浓郁^[4],在所有的苹果品种中具有一

定的代表性,因此本文将红富士苹果作为研究对象。在以往的几十年中,苹果的品质判定依赖于理化指标评价以及感官评价^[5]。然而,理化指标评价的方法耗时耗力,对检验人员的资质要求高,无法满足实际检测的需求。感官评价则是一个劳动密集型过程,客观性和可重复性较低,且不稳定性较高。如气相色谱-质谱^[6]、高效液相色谱^[7]等这些现代分析技术,也可以实现苹果的品质判定,但是这

收稿日期:2024-09-03

^{*} 基金项目:哈尔滨商业大学青年创新人才培育计划(XL0106)项目资助

些分析方法的成本高且操作过程繁琐。因此,苹果品质的无损检测对工业和消费者来说变得越来越重要^[8],实现快速检测苹果品质的传感器技术(例如电子鼻)越来越受欢迎。

电子鼻是一种模仿生物嗅觉系统的装置。电子鼻系统主要由传感器阵列和模式识别组成,它们都是嗅觉感受器和大脑的模仿者^[5]。传感器阵列用于感知气味特征信息,将化学信号转换为电信号。模式识别就是将处理后的数据信号利用气味识别算法进行分析、识别,进而判断出气味的类别。待测样本释放出含有多种有机化合物的气味,传感器阵列感知到气味时会形成特征“指纹”^[9],这些“指纹”被导入气味识别算法模型中,实现了对气体的快速检测。目前电子鼻技术已成功应用于食品工业领域^[8]。在现有研究中,电子鼻技术在水果品质检测方面有着积极应用,检测草莓^[10]、梨^[11]、桃子^[12]以及其他水果的品质^[13]。

还有部分研究尝试检测苹果的品质。Pathange等^[14]依据苹果的成熟指标对苹果的新鲜度进行无损检测,根据判别分析结果,60个样本量的分类准确率为83%。Jia等^[15]采用PEN3电子鼻检测识别接种了广青霉和黑曲霉的苹果,并采用线性判别分析、反向传播(back propagation, BP)神经网络、支持向量机和径向基函数神经网络4种模式识别方法对采集数据进行分析,其中反向传播神经网络的预测准确率可达到90%。闫转红等^[16]为了实现苹果甜度的快速无损检测,通过粒子群BP神经网络算法对100个样本量进行分类,分类准确度达到86.67%。Guo等^[17]对接种广青霉的苹果进行无损分类预测,采用多种判别分析法识别120个样本数据,分类结果良好。Qiao等^[5]将自然成熟与人工成熟的苹果作为研究对象,通过主成分分析(principal component analysis, PCA)、线性判别分析、支持向量机以及随机森林分析4种机器学习算法对120个苹果样本进行无损检测。从上述研究可以看出,目前大多数研究应用常见的气味识别算法检测苹果的品质,如主成分分析、线性判别分析、支持向量机、人工神经网络等^[18],这些单一算法存在着训练速度慢、效率低以及预报精度不足等缺点^[19]。本文首次将鲸鱼优化算法(whale optimization algorithm, WOA)应用到气味识别问题中,鲸鱼优化算法可以使BP神经网络摆脱局部最优解,提高其收敛速度,以提高BP神经网络分类识别模型的性能。

本文依据对苹果腐烂相应特征气体的研究,选择乙醇、氨气与硫化氢作为主要采集气体^[20]。利用自主设计的苹果气味感知系统采集气体数据,将气味特征数据进行PCA降维处理,提取出苹果气味的重要特征。并采用K-means、BP神经网络、鲸鱼算法优化的BP神经网络等学习算法建立苹果新鲜度预测模型,该研究为苹果新鲜度的分级提供一定的依据。

1 模式识别算法

1.1 主成分分析

主成分分析法作为统计学的分析方法之一,其广泛用于数据降维任务中^[21-22]。PCA的主要思想是利用线性变换的方式,将原始数据转换为线性不相关的变量。在该过程中,通过线性投影的方式,将高维数据映射于低维空间之中,并且使低维度上的数据方差最大化,依据数据方差的数值判断该维度上信息的重要程度,使得PCA在压缩数据维度空间的同时保留了原始数据的主要特征。

1.2 K-means 聚类

K-means聚类作为一种无监督机器学习算法,其依据距离公式以获取不同数据对象的相似度,并将相似的数据对象归于同一区域中^[23]。该算法的效果取决于区域中数据对象的相似度,相似度越高,则算法的效果越好。

步骤1)设置红富士苹果的新鲜度为聚类的类数。

步骤2)任意选择 k 个样本点作为初始聚类中心 $C = C_1, C_2, \dots, C_k$ 。

步骤3)依据欧氏距离公式 $d(x, C_j) = \sqrt{\sum_{i=1}^m (x_i - C_{ij})^2}$ 计算出其他样本至聚类中心的距离。

其中, x 为数据对象, C_j 为第 j 个聚类中心, m 是 x 的维度, x_i 和 C_{ij} 分别为 x 和 C_j 的第 i 个维度的属性值。而后经过不断的迭代,将聚类中心移至与全部对象距离最短的位置为止。

1.3 BP 神经网络

BP神经网络是现在应用最广泛的神经网络模型之一,其主要思想是依据实际输出值与期望输出值之间的关系来调整算法模型的参数,以达到降低误差的目的。BP神经网络的拓扑结构主要分为输入层、隐含层、输出层3部分^[24]。拥有多节点单隐含层的BP神经网络能够以任意精度逼近任一连续函数,所以此次研究使用隐含层数量为1的BP神经网络。在构建BP神经网络的拓扑结构时,隐含层节点数的确定决定着网络学习程度的高低。最佳的隐含层节点数可由经验公式得出:

$$M = \sqrt{u + v} + w \quad (1)$$

式中: u 为输入层节点数; v 为输出层节点数; w 为 $[1, 10]$ 区间内的常数; M 为隐含层节点数。

1.4 鲸鱼优化算法

于2016年被提出的鲸鱼优化算法是一种基于仿生学的智能优化算法,其受座头鲸捕食猎物过程的启发^[25-26]。鲸鱼的狩猎过程分为3部分,示意图如图1所示。

锁定并包围猎物。首先识别初始最优个体的位置,群内的其余个体朝最优个体移动并包围猎物,模型如下:

$$a = 2 \times (1 - t/t_{\max}) \quad (2)$$

$$A = 2a \times r_1 - a \quad (3)$$

$$E = 2 \times r_2 \quad (4)$$

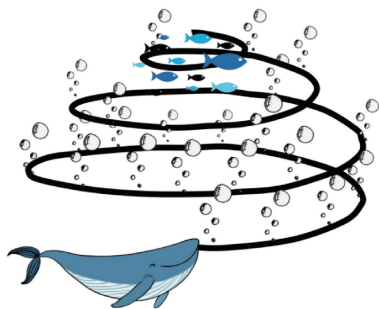


图1 鲸鱼的狩猎过程示意图

Fig.1 Schematic diagram of whale hunting process

$$D_1 = |E \times X_{\text{BEST}}(t) - X(t)| \quad (5)$$

$$X(t+1) = X_{\text{BEST}}(t) - A \times D_1 \quad (6)$$

式中: a 为收敛因子; t 为迭代次数; $t_{\max}$ 为最大迭代次数; A 、 E 代表系数向量; $X(t)$ 为鲸鱼位置向量; $X_{\text{BEST}}(t)$ 代表当前鲸鱼最佳位置向量; D_1 为鲸鱼与猎物间的距离; r_1 、 r_2 为 $[0,1]$ 内的随机数。

气泡网狩猎。鲸鱼以螺旋游动的动作靠近猎物,形成了气泡网,在该过程中,鲸鱼位置被不断更新,模型为:

$$D_2 = |X_{\text{BEST}}(t) - X(t)| \quad (7)$$

$$X(t+1) = D_2 \times e^{bl} \times \cos 2\pi l + X_{\text{BEST}}(t) \quad (8)$$

式中: b 为一个常数; l 为 $[-1,1]$ 内的一个随机数; D_2 为鲸鱼与猎物间的距离。

随机搜索猎物。当 $|A| \geq 1$ 时,鲸鱼种群依据彼此间的位置执行随机全局搜索,以寻求更合适的猎物,从而摆脱局部最优解。鲸鱼的位置更新公式为:

$$D_3 = |E \times X_{\text{RAND}}(t) - X(t)| \quad (9)$$

$$X(t+1) = X_{\text{RAND}}(t) - A \times D_3 \quad (10)$$

式中: $X_{\text{RAND}}(t)$ 为随机选取的鲸鱼位置; D_3 为鲸鱼与猎物间的距离。

1.5 鲸鱼算法优化的BP神经网络

与其他算法相比,BP神经网络拥有很多的优势,它的自学习能力与自适应能力强,但是该算法在训练的过程中容易陷入局部最优解^[27-28]。为此,引入拥有全局搜索能力的WOA优化BP神经网络的初始权值和阈值,构建出WOA-BP神经网络模型。

依据不同新鲜度的苹果所释放出气味特征的差异性,对苹果的新鲜度进行预测,进而构建苹果新鲜度的分类模型,并保存训练完成的最优模型,使得新的气味特征数据被导入时,通过分类模型能够对苹果新鲜度进行分类预测。

鲸鱼算法优化BP神经网络的步骤如下:

步骤1)输入采样数据,并对数据进行归一化处理;

步骤2)构建BP神经网络模型,确定网络的拓扑结构,初始化它的权值与阈值;

步骤3)初始化鲸鱼算法的参数,计算鲸鱼的适应度值并标记个体的最优位置;

步骤4)计算新的控制参数 a ,更新系数向量 A ,并更新个体位置;

步骤5)设置鲸鱼算法终止准则,如果满足准则,则输出全局最优位置,反之,则返回步骤3);

步骤6)获得最优的权值和阈值,将最优参数赋予给BP神经网络,从而输出WOA-BP算法模型。

2 材料与方法

2.1 实验材料

1)苹果选材

本文选取不同新鲜度的红富士苹果作为研究对象。苹果的新鲜度分为3个等级,分别为新鲜、次新鲜和腐烂。关于苹果新鲜度的鉴别方法采用介电特性法^[4],该方法发展较早,技术相对成熟。将采购的所有大小均匀的苹果样本进行介电常数测定后,对比数值并对样本进行新鲜度分类,得到不同新鲜度的苹果样本各60个,共计180个。

2)主要仪器和设备

实验所用的电子鼻系统由5颗MOS气敏传感器组成的阵列感知苹果气味,各气敏传感器的特性如表1所示。实验设备还含有驱动传感器阵列的核心控制器和显示采集数据的上位机。

表1 气敏传感器型号及参数

Table 1 Model and parameters of gas sensor

编号	传感器型号	响应气体	检测范围 ($\times 10^{-6}$)
S1	TGS2620	乙醇、氢气等	50~5 000
S2	TGS2600	氢气、乙醇、异丁烷等	1~30
S3	TGS2602	VOC、氨气、硫化氢等	1~30
S4	TGS2603	三甲胺、甲硫醇、乙醇等	1~10
S5	TGS2610	异丁烷、液化石油气、乙醇等	500~10 000

2.2 实验方法

将传感器阵列、核心控制器与单个苹果样本平稳地置于气体反应室中,其中长宽高为17 cm $\times$ 17 cm $\times$ 12 cm的透明亚克力箱子作为气体反应室。整个气味采集实验过程在超净台内完成,以确保空气的新鲜度,排除干扰气体存在的可能性。实验现场如图2所示。



图2 实验实景图

Fig.2 Experimental scene picture

电子鼻系统中传感器阵列与核心控制器相连接以便采集气味数据,将化学信号转换为电信号。核心控制器通过串口与上位机相连接,将采集的数据传输给上位机。苹果样本气味数据的具体采集步骤如下:

步骤1)将电路连通并按下超净台的通风按钮,使传感器预热处理0.5 h,直至传感器阵列的输出电压值处于稳定状态;

步骤2)从180个苹果样本中依次取出单个样本放入气体反应室中,并扣紧反应室的盖子,关闭超净台的通风按钮;

步骤3)静置1 h后,利用上位机保存传感器阵列的响应数据,而后取出样本并打开超净台的通风按钮,直至传感器阵列的输出电压值再次达到稳态;

步骤4)返回步骤2)、3),直到180个苹果的气味数据全部采集完毕为止。

3 实验结果

3.1 特征值提取

为研究传感器阵列对红富士苹果气味的感知程度并获得有效特征值,根据传感器的基本测试电路,得出输出电压 V_{OUT} 的求解公式^[20]:

$$V_{OUT} = \frac{R_L}{R_L + R_S} \times V_C \quad (11)$$

式中: R_L 为负载电阻; V_C 为传感器的供电电压; R_S 为传感器素子的电阻。从而得到1 h内输出电压值随采集点数变化的气味数据采集曲线如图3所示。其中图3(a)为新鲜苹果的气味采集数据曲线,图3(b)为次新鲜苹果的气味采集数据曲线,图3(c)为腐烂苹果的气味采集数据曲线。由于传感器采样的时间间隔为3 s,因此1 h内传感器阵列的采样点数为1 200个。

3.2 数据预处理

由于传感器本身的局限性或外界干扰因素,导致传感器在实验中所采集的数据有误差,为了减少误差并保留数据的主要特征,对数据进行 Savitzky-Golay 卷积平滑(窗口宽度设定为5,多项式次数设定为2)处理^[29],过滤数据中的噪声。Savitzky-Golay 卷积平滑是一种基于移动平均平滑改进的信号处理技术,其原理是在信号上执行滑动窗口的线性拟合,利用对窗口中的数据进行多项式拟合来计算窗口中心的估计值,以此完成信号的平滑处理。则有:

$$x_{k,smooth} = \frac{1}{H} \sum_{i=-w}^{+w} x_{k+i} h_i \quad (12)$$

式中: H 为归一化因子; x_k 为窗口中心点。

进行光滑处理后的曲线如图4所示,其中图4(a)为过滤后新鲜苹果的气味采集数据曲线,图4(b)为过滤后次新鲜苹果的气味采集数据曲线,图4(c)为过滤后腐烂苹果的气味采集数据曲线。对比图4(a)~(c)可以看出,每幅图的传感器响应曲线的趋势存在一定程度的差异性。新鲜苹果的气味采集数据曲线的倾斜度最小,次新鲜苹果的

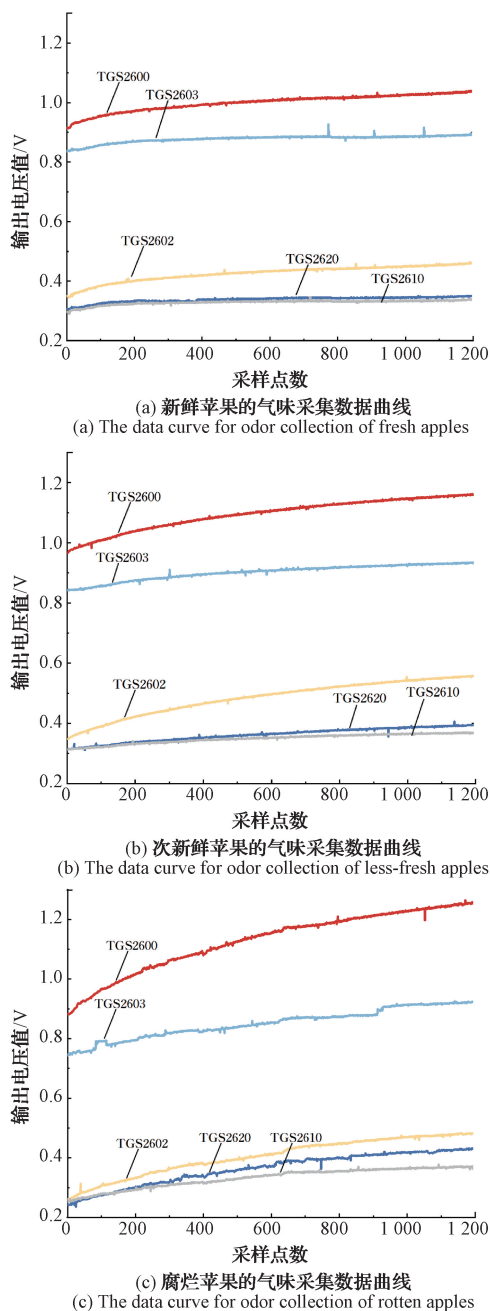


图3 气味数据采集曲线

Fig. 3 The data curve for odor collection

气味采集数据曲线次之,腐烂苹果的倾斜度最大,这表明随着苹果新鲜度等级的变化,苹果所散发的气体浓度值越大。因此苹果的新鲜度与其散发出气味浓度之间存在着特定的关系,可将其作为识别苹果新鲜度的指标。

3.3 特征数据集构建

从气味采集数据曲线可以看出,曲线的极差值与斜率各有差异,因此提取曲线的极差值与斜率值作为苹果气味特征。定义曲线起始点处的输出电压值为 V_0 ,定义如下提取的特征值。

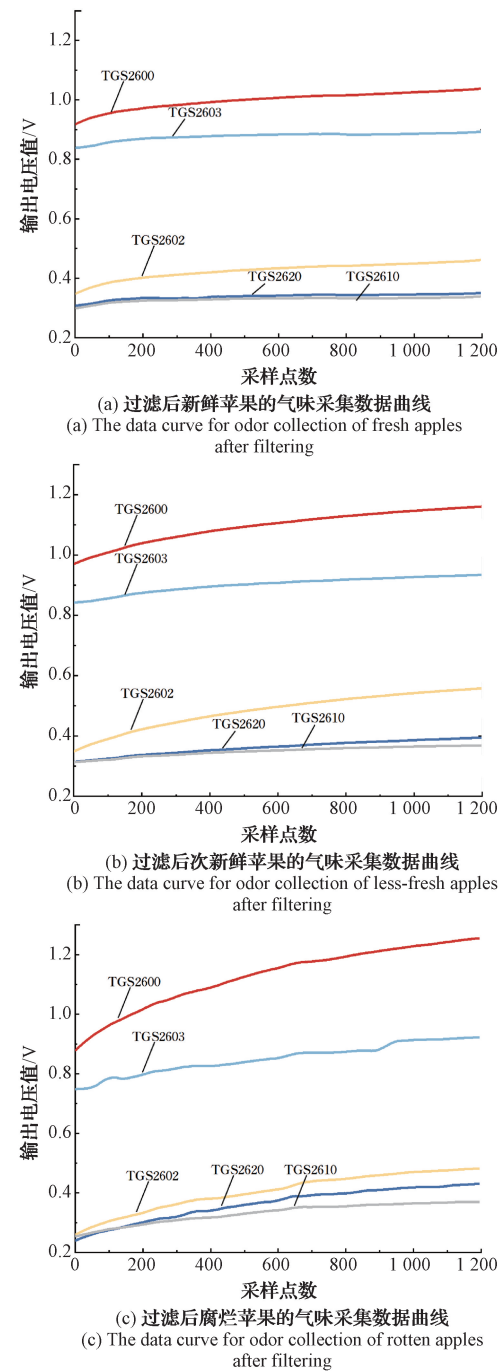


图4 过滤后的气味数据采集曲线
Fig. 4 Filtered odor collection data curve

- 1)斜率值,提取曲线起始点之后的5个数据点对应的输出电压值为 $V_1 \sim V_5$,依次计算5个数据点与起始点连线的斜率 $[(V_x - V_0)/\Delta X]$ 作为特征值,定义为 $K_1, K_2, \dots, K_5$ 。
- 2)极差值,提取曲线的极小值 $V_{\min}$ 和极大值 $V_{\max}$,计算极大值与极小值的差值 $[V_{\max} - V_{\min}]$ 作为特征值,定义为 B 。

从每颗传感器所采集的数据中提取到6个特征值,5颗传感器总共提取到30个特征值,部分特征数据如表2所示。将所有的特征值按顺序排列构成特征向量,构成一个由180个特征值组成的数据集合。特征数据排列示意图如图5所示。

表2 部分气味特征数据
Table 2 Partial odor characteristic data

传感器 型号	K_1 /mV	K_2 /mV	K_3 /mV	K_4 /mV	K_5 /mV	B/V
TGS2620	0.350	0.305	0.267	0.251	0.239	0.191
TGS2600	0.833	0.687	0.600	0.525	0.493	0.376
TGS2602	0.455	0.365	0.341	0.303	0.271	0.222
TGS2603	0.349	0.237	0.228	0.192	0.178	0.175
TGS2610	0.241	0.199	0.182	0.158	0.152	0.116

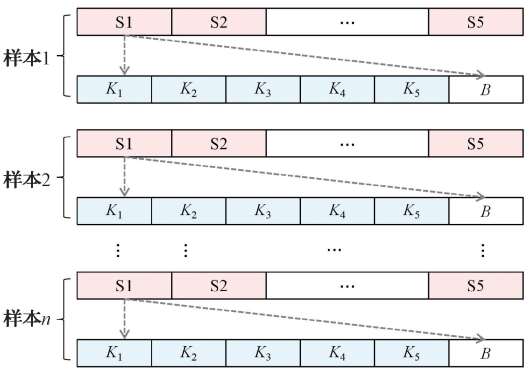


图5 特征数据排列示意图
Fig. 5 Schematic diagram of characteristic data arrangement

3.4 PCA 分析

该数据集合一共包含 180×30 个数据,从该数据集中提取出 120×30 个数据集合,通过PCA分析对提取出的数据集合展开降维处理。经过PCA分析处理后,特征数据由特征向量方向上的信息量进行表征,特征值与坐标维度上的方差相等,较大的方差值更能表示原始数据。特征向量的方差贡献率表示对应维度下所蕴含原始信息量的比例,其值为特征值与所有特征值之和的比值。PCA分析的主成分示意图如图6所示,可以看出,数据集合经过PCA分析后全部特征值的比例,并显示出所有主成分的累计贡献率。由图6可得特征值比例的最大值达到65.89%,占比较大的特征值为前5个,其主成分累计贡献率为98.05%。

将处理的数据集可视化,分析结果如图7所示。图7(a)是前2个主成分的二维PCA分析,可知,第1主成分的贡献率为65.9%,而第2主成分的贡献率为22.9%,贡献率之和为88.8%,说明这2个主成分包含着大部分原始信息。由图7(a)还看出不同新鲜度的分布趋势,腐烂苹果

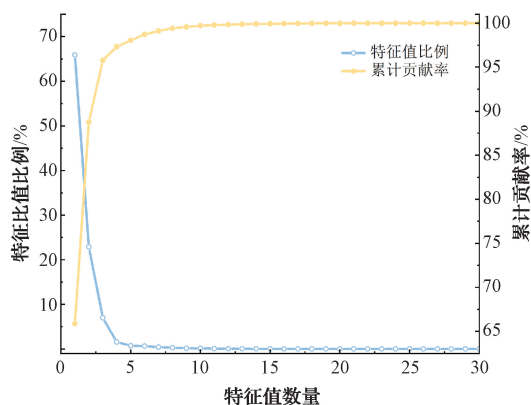
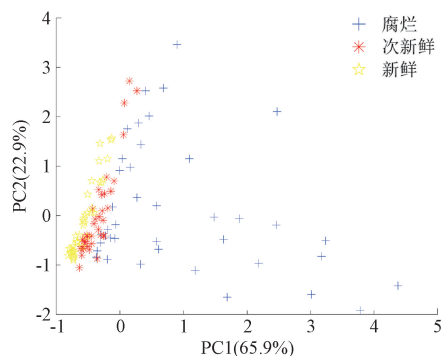


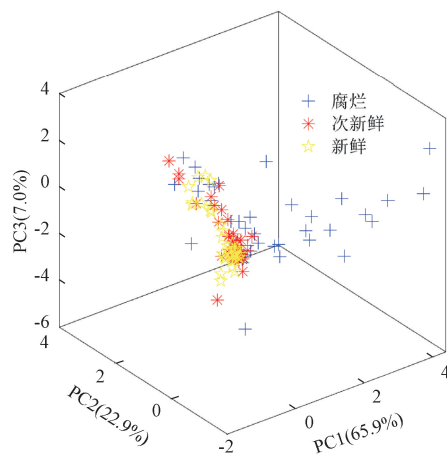
图6 PCA分析的主成分示意图

Fig. 6 Schematic diagram of principal components of PCA analysis



(a) 二维PCA分析

(a) Two-dimensional PCA analysis diagram



(b) 三维PCA分析

(b) Three-dimensional PCA analysis diagram

图7 PCA分析

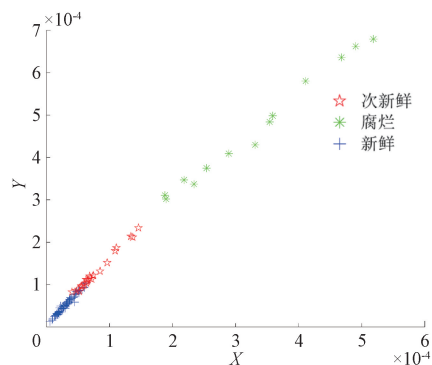
Fig. 7 PCA analysis diagram

样本分布范围更广,这表示样本特征的特征方差与离散性都比较大。图7(b)为前3个主成分的三维PCA分析。从图7(b)可以看出,不同新鲜度苹果样本的区分度比图7(a)更佳,但一些样本仍存在无法区分的情况。由图7可

知,利用PCA分析难以对样本数据直接分类,然而利用PCA降维后不同新鲜度样本的离散性存在一定的差异,这表明利用聚类算法对苹果样本的新鲜度进行区分是可行的。

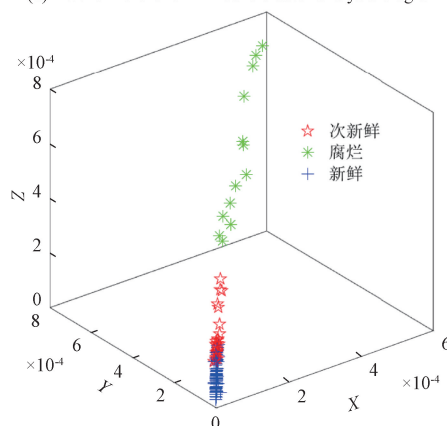
3.5 K-means 聚类分析

既然利用聚类的方法对样本的新鲜度进行分类具有一定的可行性,所以通过K-means聚类算法将样本数据进行分类处理。将算法的 k 值设置为3,对应样本的3个新鲜度等级。由于样本特征数据的总数为120个,其中每个新鲜度等级各40个,所以理想状态下每个聚类都应含有40个同一新鲜度等级的样本数据。二维与三维的K-means聚类分析如图8所示,算法的聚类精度为61.67%。图8中样本数据呈长条状分布,腐烂样本明显区别于其余样本,次新鲜样本与新鲜样本存在部分交错重叠的情况,这说明采用K-means聚类算法对样本原始特征的选择性不够强。



(a) 二维K-means聚类分析

(a) Two-dimensional K-means cluster analysis diagram



(b) 三维K-means聚类分析

(b) Three-dimensional K-means cluster analysis diagram

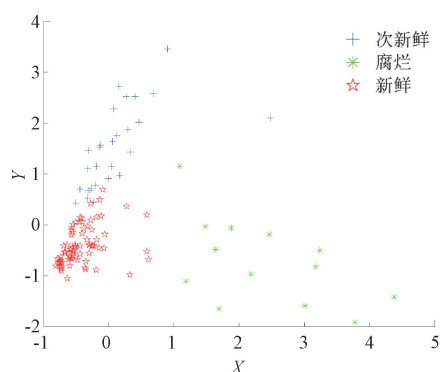
图8 K-means 聚类分析

Fig. 8 K-means cluster analysis diagram

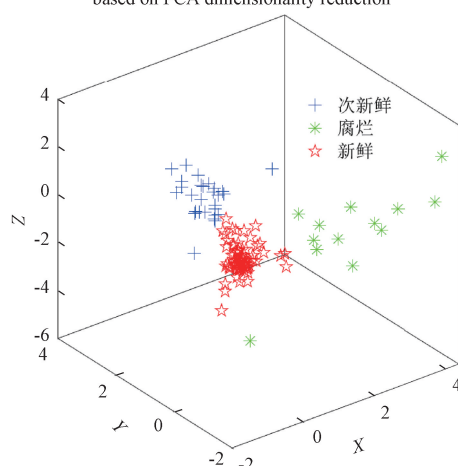
将经过PCA降维处理后的特征数据进行K-means聚类分析,得到的二维与三维分析如图9所示。图9中样本数据并未呈现出明显的交错重叠现象,且不同新鲜度样本的分布区域清晰明了,这表明经过PCA降维处理后的样

本特征更有效。基于PCA降维处理的K-means聚类分析的聚类精度达到了70.67%。

虽然图9中不同新鲜度样本的区分度较高,分布范围清晰可见,但是每一新鲜度样本的聚类总数存在明显差异,比如腐烂样本的聚类数量仅有13个,其值远低于理想状态下的40个样本,算法的聚类精度被直接拉低。



(a) 基于PCA降维的二维K-means聚类分析
(a) Two-dimensional K-means cluster analysis diagram based on PCA dimensionality reduction



(b) 基于PCA降维的三维K-means聚类分析
(b) Three-dimensional K-means cluster analysis diagram based on PCA dimensionality reduction

图9 基于PCA降维的K-means聚类分析

Fig.9 K-means cluster analysis diagram based on PCA dimensionality reduction

3.6 WOA-BP神经网络分析

为了进一步研究样本数据之间的关系以实现分类识别,采用WOA-BP神经网络算法对气味特征数据集进行分析处理。

1) 数据归一化处理

将 180×30 个苹果气味特征数据集作为WOA-BP神经网络的输入,因而BP神经网络的输入层节点数为30个。由于输入层的各个参数数量级差距过大会影响算法的运行速率和精度,所以对输入参数进行归一化处理^[16]:

$$Y = \frac{Y_n - Y_{\min}}{Y_{\max} - Y_{\min}} \quad (13)$$

式中: Y_n 为第 n 个特征数据; $Y_{\min}$ 为 Y_n 所在数据列的最小值; $Y_{\max}$ 为 Y_n 所在数据列的最大值; Y 为 Y_n 的归一化值。

2) WOA-BP神经网络结构的设计

从苹果气味特征数据集中随机提取出 120×30 个数据集作为训练集,其余 60×30 个数据集作为测试集。将苹果的3个新鲜度等级作为输出层,即输出层节点数为3。根据式(1)可以得到隐含层节点数范围为 $[7, 16]$,经过数次试验后得到最佳隐含层节点数为15。将学习速率设为0.01,训练次数设为1000,目标误差为0.000001,K交叉验证的 K 值为10。对于WOA算法,初始种群规模为20,且最大迭代次数设为50,构建基于WOA-BP神经网络算法的分类识别模型。

3) 评估指标

为验证本文所提方法的有效性,选取召回率、精确率、 F 值、总体准确率以及ROC曲线作为评估指标对苹果新鲜度的分类模型进行性能评估,各个评估指标如式(14)~(17)所示。本文利用混淆矩阵辅助计算评估指标,混淆矩阵如表3所示。

$$\text{召回率} = \frac{TP}{TP + FN} \quad (14)$$

$$\text{精确率} = \frac{TP}{TP + FP} \quad (15)$$

$$F \text{ 值} = \frac{2PR \cdot RR}{PR + RR} \quad (16)$$

$$\text{总体准确率} = \frac{\sum_{i=1}^N TP_i}{N} \quad (17)$$

式中: N 为样本总数。

表3 混淆矩阵

Table 3 Confusion matrix

真实/预测	正例(P)	反例(N)
正例(P)	TP	FP
反例(N)	FN	TN

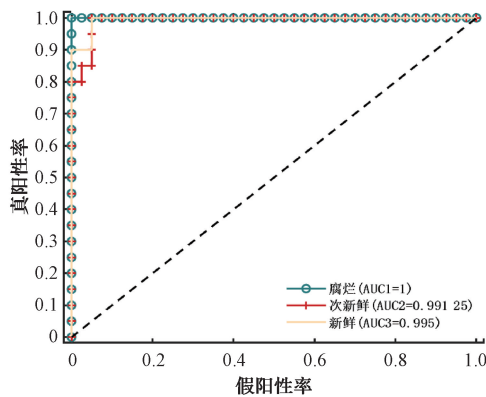
将经过PCA降维处理的气味特征数据和原始特征数据作为输入分别导入BP神经网络模型与WOA-BP神经网络模型中,得到不同分类模型的评估指标,如表4所示。基于PCA降维特征的神经网络模型得到的总体准确率分别为0.8167和0.75,不如基于原始特征的神经网络模型。从式(16)可以看出, F 值是由召回率与精确率共同得出的,所以可以选取 F 值作为衡量模型精度的指标。由表4可知,基于原始特征的WOA-BP神经网络模型的3个 F 值都在0.9以上,且明显高于其余模型,说明基于原始特征的WOA-BP神经网络模型对苹果新鲜度的分类识别效果更好。

ROC曲线能够直观地反映出分类模型在不同阈值下的性能表现。一般来说,ROC曲线距离坐标轴左上角越近,分类模型的性能越好。AUC值被定义为ROC曲线下

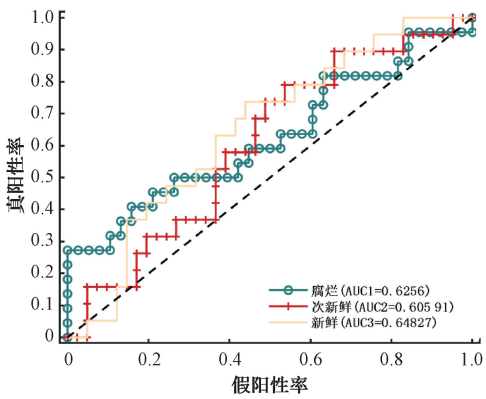
表 4 评估指标
Table 4 Evaluation indicators

分类模型	新鲜度	召回率	精确率	F 值	总体 准确率
PCA+BP	腐烂	0.615 4	0.941 2	0.744 2	0.816 7
	次新鲜	0.789 5	0.681 8	0.731 7	
	新鲜	0.692 3	0.857 1	0.765 9	
BP	腐烂	0.703 7	0.950 0	0.808 5	0.850 0
	次新鲜	0.818 2	0.782 6	0.8	
	新鲜	0.700 0	0.823 5	0.756 7	
PCA+ WOA-BP	腐烂	0.576 9	0.789 5	0.666 7	0.750 0
	次新鲜	0.695 7	0.666 7	0.680 9	
	新鲜	0.538 5	0.823 5	0.651 2	
WOA-BP	腐烂	0.869 6	1.000 0	0.930 3	0.950 0
	次新鲜	0.950 0	0.904 8	0.926 8	
	新鲜	0.900 0	0.947 4	0.923 1	
K-means	腐烂	0.406 3	0.325	0.361 1	0.616 7
	次新鲜	0.440 0	0.55	0.488 9	
	新鲜	0.464 3	0.975	0.629 0	
PCA+ K-means	腐烂	0.619 0	0.325	0.426 2	0.706 7
	次新鲜	0.539 7	0.85	0.660 2	
	新鲜	0.535 2	0.95	0.684 7	

方区域的面积,其取值范围在 $[0,1]$ 之间。AUC 值越大,则分类模型性能越好。ROC 曲线如图 10 所示,图 10(a)为 WOA-BP 神经网络模型的 ROC 曲线,可以看出,3 个新鲜度等级的 AUC 值均在 0.99 以上。图 10(b)为 BP 神经网络模型的 ROC 曲线,其中 3 个新鲜度等级的 AUC 值均在 0.6 以上。结果表明,通过 WOA 优化 BP 神经网络的初始权值和阈值,可以提升数据分类效果。由此可知,WOA-BP 神经网络模型的 ROC 曲线面积较大,分类效果较好。



(a) WOA-BP神经网络模型的ROC曲线
(a) ROC curve of WOA-BP neural network model

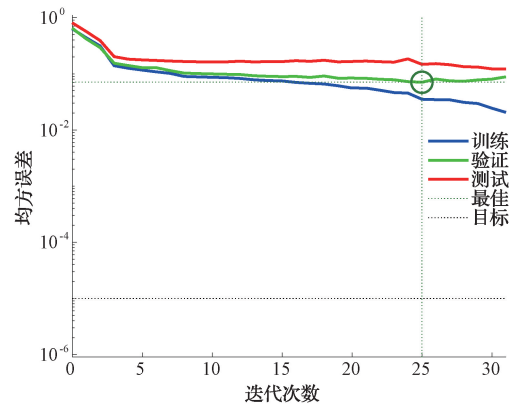


(b) BP神经网络模型的ROC曲线
(b) ROC curve of BP neural network model

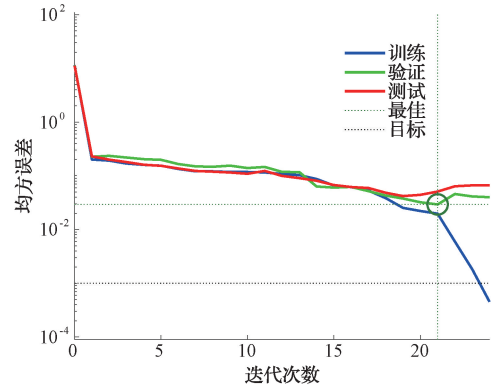
图 10 ROC 曲线
Fig. 10 ROC curve

4) 预测结果

利用 MATLAB 2021 平台进行模型训练的训练结果如图 11 所示。图 11(a)为 BP 神经网络模型训练结果,图 11(b)为 WOA-BP 神经网络模型训练结果。由图 11(a)可知,BP 神经网络经 25 次迭代后均方误差(MSE)达到最小值 0.070 68,而在图 11(b)中 WOA 算法



(a) BP神经网络模型训练结果
(a) Training results of BP neural network model



(b) WOA-BP神经网络模型训练结果
(b) Training results of WOA-BP neural network model

图 11 模型训练结果

Fig. 11 Training results of model

优化的BP神经网络经21次迭代后均方误差达到最小值0.029347。表明,WOA算法优化的BP神经网络在训练效率与误差方面比传统的BP神经网络表现更优。

为方便算法模型的训练,利用数字的形式来表示苹果新鲜度等级的输出,表示形式如表5所示。苹果新鲜度的分类识别结果如图12所示,其中图12(a)为BP神经网络模型的分类识别结果,图12(b)为WOA-BP神经网络模型的分类识别结果。从图12可以看出,BP神经网络的第6、18、23、25、32、34、37、48、53组样本未被正确预测苹果新鲜度,对应测试集的准确率为85%。WOA-BP神经网络的第27、31、51组样本未被正确预测苹果新鲜度,对应测试集的准确率达到了95%。结果表明,WOA-BP神经网络的分类识别精度更高。

表5 苹果新鲜度等级的输出表示形式

Table 5 Output representation of apple freshness rating

苹果新鲜度	十进制	二进制
腐烂	1	100
次新鲜	2	010
新鲜	3	001

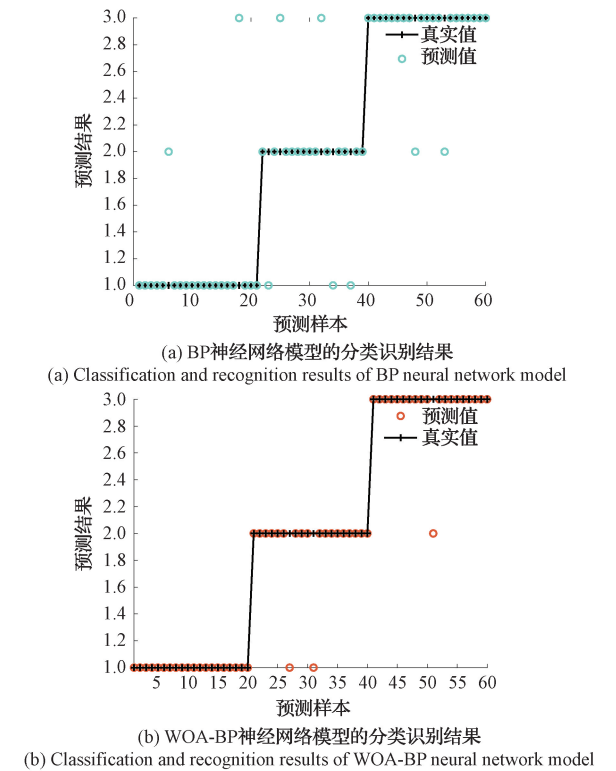


图12 苹果新鲜度的分类识别结果

Fig. 12 Classification and recognition results of apple freshness

4 结论

本文将主成分分析、K-means 聚类、BP 神经网络、鲸

鱼算法等多种算法模型应用于苹果新鲜度的分类识别中。实验结果表明,经过 PCA 降维处理的苹果气味特征数据可以提高 K-means 聚类算法的聚类精度,使聚类精度由 61.67% 提高到 70.67%。鲸鱼算法优化的 BP 神经网络算法对原始特征数据进行分类识别,识别准确率可达 95%,识别效果优于 BP 神经网络算法。本文应用电子鼻系统对红富士苹果的新鲜度进行分类识别,克服了感官评价方法的不确定性与仪器分析方法的繁琐、费用较高等问题,使识别检测过程高效便捷,应用前景较好。

参考文献

[1] ZHU D, REN X, WEI L, et al. Collaborative analysis on difference of apple fruits flavour using electronic nose and electronic tongue [J]. Scientia Horticulturae, 2020, 260: 108879.

[2] 马勉娣,曾厅余,鲁兴凯,等. 云南昭通与环渤海湾等3个产区红富士苹果品质的比较[J]. 中国果树, 2024(5): 33-38.

MA M D, ZENG T Y, LU X K, et al. Comparison of 'Red Fuji' apple quality between Zhaotong Yunnan and three production regions including around Bohai Bay[J]. China Fruits, 2024(5): 33-38.

[3] 陈学森,毛志泉,王志刚,等. 持续多代芽变选种及其芽变机理揭开'红富士'在我国苹果产业独占鳌头的谜底[J]. 中国果树, 2020(3): 7-11, 148.

CHEN X S, MAO ZH Q, WANG ZH G, et al. Continuous multigenerational sports selection and its mechanism reveals the mystery of 'Red Fuji' in China's apple industry[J]. China Fruits, 2020(3): 7-11, 148.

[4] 杨伟震. 基于气体传感器阵列的苹果鲜度预测技术研究[D]. 太原: 中北大学, 2023.

YANG W ZH. Research on prediction technology of apple freshness based on sensor array[D]. Taiyuan: North University of China, 2019.

[5] QIAO J, SU G, LIU C, et al. Study on the application of electronic nose technology in the detection for the artificial ripening of crab apples[J]. Horticulturae, 2022, 8(5): 386.

[6] FUJIOKA, K. Evaluating the non-invasive measurement of apple aroma using electronic nose device through comparison with direct mass spectrometry, sugar content, and ripeness measurements[J]. Sensors, 2024, 24: 3114.

[7] MASTELLO R B, CAPOBIANGO M, CHIN S T, et al. Identification of odour-active compounds of pasteurised orange juice using multidimensional gas chromatography techniques [J]. Food Research International, 2015, 75: 281-288.

- [8] CARMONA E N, SBERVEGLIERI V, PONZONI A, et al. Detection of food and skin pathogen microbiota by means of an electronic nose based on metal oxide chemiresistors [J]. *Sensors and Actuators B: Chemical*, 2017, 238: 1224-1230.
- [9] 杨志蒙, 黄波, 赵永礼. 基于气味检测的茶叶种类识别方法研究[J]. *农业装备与车辆工程*, 2024, 62(2): 34-39.
- YANG ZH M, HUANG B, ZHAO Y L. Research on tea category identification method based on odor detection [J]. *Agricultural Equipment & Vehicle Engineering*, 2024, 62(2): 34-39.
- [10] XING M, SUN K, LIU Q, et al. Development of novel electronic nose applied for strawberry freshness detection during storage[J]. *International Journal of Food Engineering*, 2018, 14(7-8): 20180111.
- [11] ALI A, MANSOL A S, KHAN A A, et al. Electronic nose as a tool for early detection of diseases and quality monitoring in fresh postharvest produce: A comprehensive review[J]. *Comprehensive Reviews in Food Science and Food Safety*, 2023, 22(3): 2408-2432.
- [12] XIN R, LIU X, WEI C, et al. E-nose and GC-MS reveal a difference in the volatile profiles of white- and red-fleshed peach fruit[J]. *Sensors*, 2018, 18: 765.
- [13] WEN T, ZHENG L, DONG S, et al. Rapid detection and classification of citrus fruits infestation by *Bactrocera dorsalis* (Hendel) based on electronic nose[J]. *Postharvest Biology and Technology*, 2019, 147: 156-165.
- [14] PATHANGE L P, MALLIKARJUNAN P, MARINI R P, et al. Non-destructive evaluation of apple maturity using an electronic nose system[J]. *Journal of Food Engineering*, 2006, 77(4): 1018-1023.
- [15] JIA W, LIANG G, TIAN H, et al. Electronic nose-based technique for rapid detection and recognition of moldy apples[J]. *Sensors*, 2019, 19(7): 1526.
- [16] 闫转红, 王伟. 基于气体传感器阵列的苹果甜度识别技术研究[J]. *国外电子测量技术*, 2021, 40(10): 71-76.
- YAN ZH H, WANG W. Research on apple sweetness recognition technology based on gas sensor array[J]. *Foreign Electronic Measurement Technology*, 2021, 40(10): 71-76.
- [17] GUO Z, GUO C, CHEN Q, et al. Classification for *Penicillium expansum* spoilage and defect in apples by electronic nose combined with chemometrics [J]. *Sensors*, 2020, 20(7): 2130.
- [18] 刘云刚, 王伟. 基于 SFLA 优化的 BP 神经网络苹果鲜度气味识别系统[J]. *传感器与微系统*, 2020, 39(8): 96-99, 106.
- LIU Y G, WANG W. Apple fresh odor recognition system based on SFLA optimized BP neural network[J]. *Transducer and Microsystem Technologies*, 2020, 39(8): 96-99, 106.
- [19] TANG Y, XU K, ZHAO B, et al. A novel electronic nose for the detection and classification of pesticide residue on apples[J]. *RSC Advances*, 2021, 11(34): 20874-83.
- [20] 熊忆舟. 基于仿生嗅觉神经网络的电子鼻食品气味识别算法研究[D]. 杭州: 浙江大学, 2022.
- XIONG Y ZH. Research on food odor recognition algorithm of electronic nose based on bioinspired olfactory neural network [D]. Hangzhou: Zhejiang University, 2022.
- [21] 侯琳杰, 解维坤, 陈世博, 等. 一种基于神经网络的 THD 参数快速估计算法[J]. *仪器仪表学报*, 2023, 44(2): 101-109.
- HOU L J, XIE W K, CHEN SH B, et al. Fast estimation algorithm of THD based on artificial neural network[J]. *Chinese Journal of Scientific Instrument*, 2023, 44(2): 101-109.
- [22] INAN O, UZER M S. A method of classification performance improvement via a strategy of clustering-based data elimination integrated with k-fold cross-validation [J]. *Arabian Journal for Science and Engineering*, 2021, 46(2): 1199-1212.
- [23] LIU H. Optimal selection of control parameters for automatic machining based on BP neural network[J]. *Energy Reports*, 2022, 8: 7016-7024.
- [24] SUN Z, SONG Z, YUE B, et al. Inversion model for surrounding rock parameters based on back propagation neural networks improved by stochastic gradient descent algorithm[J]. *Tunnel Construction*, 2023, 43(12): 2066.
- [25] 张家刘, 李正权. 基于改进鲸鱼算法优化模糊控制的开关电源[J]. *电子测量技术*, 2024, 47(1): 87-92.
- ZHANG J L, LI ZH Q. Optimization of fuzzy control for switching power supply based on improved whale algorithm [J]. *Electronic Measurement Technology*, 2024, 47(1): 87-92.
- [26] LIANG Z, HAN Q, ZHANG T, et al. Nonlinearity compensation of magneto-optic fiber current sensors based on WOA-BP neural network[J]. *IEEE Sensors Journal*, 2022, 22(20): 19378-19383.
- [27] 王力, 张露露. 基于 IHHO-BP 神经网络的模拟电路故障诊断[J]. *电子测量与仪器学报*, 2024, 38(5): 238-248.

- WANG L, ZHANG L L. Fault diagnosis of analog circuit based on IHHO-BP neural network[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(5):238-248.
- [28] 李鹏. 基于改进 PSO-BP 算法的机器人目标位姿识别方法[J]. 国外电子测量技术, 2023, 42(1):7-12.
- LI P. Robot target pose recognition method based on improved PSO-BP algorithm[J]. Foreign Electronic Measurement Technology, 2023, 42(1):7-12.
- [29] 顾思思, 常世新, 韩明敏, 等. 毫米波雷达数据驱动的桥梁挠度预测[J]. 科学技术与工程, 2023, 23(11): 4874-4880.
- GU S S, CHANG SH X, HAN M M, et al. Bridge deflection prediction driven by millimeter wave radar

data[J]. Science Technology and Engineering, 2023, 23(11): 4874-4880.

作者简介

杨明丽, 硕士研究生, 主要研究方向为气味识别、烟羽追踪等。

E-mail: yml18846026592@163.com

纠海峰(通信作者), 博士, 讲师, 主要研究方向为机器人控制及路径规划技术等。

E-mail: jiuhaifeng@hotmail.com

邓薇, 博士, 讲师, 主要研究方向为智能决策分析算法等。

E-mail: dengwei1110@126.com