

黄琼, 杨红云, 万颖. 基于特征数据的水稻种子分类识别方法[J]. 江苏农业学报, 2021, 37(1): 8-15.
doi: 10.3969/j.issn.1000-4440.2021.01.002

基于特征数据的水稻种子分类识别方法

黄琼¹, 杨红云², 万颖¹

(1. 江西农业大学计算机与信息工程学院, 江西 南昌 330045; 2. 江西农业大学软件学院, 江西 南昌 330045)

摘要: 针对水稻种子相似度高、识别困难等问题, 提出一种线性判别分析(Linear discriminant analysis, LDA)和贝叶斯分类(Bayes)相结合的分类识别方法, 以提高水稻种子分类识别速度和识别准确率。通过对4类水稻种子(楚梗7号、马坝油粘、玉杨糯、玉针香)的图像进行裁剪和分割等预处理操作, 提取出水稻种子图像的颜色特征、几何特征和纹理特征。利用线性判别分析、主成分分析、因子分析和局部线性嵌入对特征数据进行分析降维, 并分别选择Bayes、K-邻近、支持向量机、多层感知机分类器对原始特征数据和降维数据进行分类识别研究。为提高模型泛化能力, 通过图像增强技术对稻种原始数据集进行样本扩充, 利用图像增强技术模拟多种环境对水稻种子图片数据集进行增强处理, 结果显示, 基于数据增强后的LDA_Bayes模型运行时间为0.019 s, 识别准确率为99.4%。与其他模型比较, 该模型具有更强的鲁棒性和适用性, 能高效地分类识别不同环境下的水稻种子, 可为水稻种子分类识别提供一种新方法。

关键词: 水稻种子; 线性判别; Bayes; 分类识别

中图分类号: S126; S511 **文献标识码:** A **文章编号:** 1000-4440(2021)01-0008-08

Classification and recognition method of rice seeds based on feature data

HUANG Qiong¹, YANG Hong-yun², WAN Ying¹

(1. College of Computer and Information Engineering, Jiangxi Agricultural University, Nanchang 330045, China; 2. College of Software, Jiangxi Agricultural University, Nanchang 330045, China)

Abstract: Aiming at the problems of high similarity and difficult identification of rice seeds, a classification and identification method which combined linear discriminant analysis (LDA) and Bayesian classification (Bayes) was proposed to improve the identification speed and accuracy of rice seed classification. By performing pre-processing operations such as cropping and segmentation on the images of four types of rice seeds (Chujing No.7, Maba Younian, Yuyangnuo and Yuzhenxiang), the color characteristics, geometric and texture features of rice seed images were extracted. Linear discriminant analysis, principal component analysis, factor analysis and locally linear embedding were used to analyze and reduce the dimensionality of the feature data, and Bayes, K-nearest neighbors, support vector machine, multilayer perceptron classifiers were selected respectively to conduct classification and identification research on the original feature data and dimensionality reduced data. To improve the generalization ability of the model, image enhancement technology was used to extend the original data set of rice seed image samples and to simulate multiple environments to enhance the image data set of rice seeds. The results showed that the running time of LDA_Bayes model based on data enhancement was 0.019 s and the

recognition accuracy rate was 99.4%. Compared with other models, LDA_Bayes model shows stronger robustness and applicability, it can classify and recognize rice seeds in different environments efficiently and provide a new method for rice seed classification and identification.

Key words: rice seeds; linear discrimination;

收稿日期: 2020-06-29

基金项目: 国家自然科学基金项目(61562039)

作者简介: 黄琼(1997-), 女, 江西赣州人, 硕士研究生, 主要从事图形图像处理研究。(E-mail) 406274397@qq.com

通讯作者: 杨红云, (E-mail) nc_yhy@163.com

Bayes; classification and recognition

水稻种子是农业生产的基本生产资料,不同品种的水稻在抗病性、生长环境及产量收益方面存在差异性,种子纯度是衡量水稻种子质量的主要指标,而水稻种子的正确分类识别可以有效保证水稻种子的纯度。目前,水稻种子分类方法如形态鉴定法、苯酚染色法等都是靠人工识别,这些方法存在耗时长、精度低、易受人的主观影响等弊端,识别速度无法满足实际要求。人工识别水稻品种比较困难,每年会因品种识别错误和纯度差造成巨大的经济损失。为了保证水稻种子纯度、减少其他种子混杂、提高选种育种的质量、保证农业正常生产,需要解决如何快速准确地分类识别日益增长的水稻品种的问题。机器视觉具有效率高、耗时短、客观性强等优点,利用机器视觉对水稻种子进行识别可以弥补人工识别的不足,是水稻种子品种识别的可行方法。

目前,国内外对水稻种子品种的识别研究主要通过仿生鼻、光谱仪、图像^[1-2]来获取水稻种子的特征参数。徐赛等^[3]利用 PEN 电子鼻采集水稻种子样品气味数据结合 BP(反向传播)神经网络对不同水稻种子进行分类识别,分类准确率为 90%。李辉等^[4]利用激光诱导击穿光谱数据结合 BP 神经网络对水稻品种进行识别,识别率可达 92.4%。Hai^[5]提取水稻种子的高光谱特征并与几何特征结合,利用支持向量机(SVM)和随机森林(RF)对数据进行分类,分类精度为 84%。Fabiya 等^[6]从高分辨率 RGB(红-绿-蓝)图像和高光谱图像中提取水稻种子的图像特征,利用 RF 分类器对特征数据进行分类,可以有效提高种子纯度。邓小琴等^[7]融合水稻种子的光谱特征、纹理特征和几何特征,建立 PLSDA 分类模型,训练精度为 99.33%。以上方法通过 PEN 电子鼻、高光谱和图像提取的特征数据相结合,利用分类器对水稻种子进行识别,分类模型识别率较高,但 PEN 电子鼻和光谱仪的成本过高,不利于推广。利用图像获取水稻种子参数的方法容易实现,便于推广,李冬^[8]通过对水稻种子的几何特征进行主成分分析(PCA),采用余弦相似度对种子进行鉴别,准确率为 88%。郑宇^[9]采用检测算法自动提取水稻种子的 13 个几何特征,利用 Bayes 判别法建立水稻种子分类模型,识别正确率可达 90%。上述方法只单一地提取了水稻种子图像的几何特征,忽略了如颜色、纹理等其他特征参数对分类

带来的影响。李秀吴等^[10]将提取的水稻种子颜色和形状特征融合,利用 SVM 对水稻种子进行分类,准确率可达 99.5%,运行时间为 0.165 s。黄星奕等^[11]提出把水稻种子图像的 shape 特征和颜色特征结合,通过贝叶斯分类器,识别正确率可达 88.3%。Hong 等^[12]通过提取水稻种子图像的全局特征和局部特征,采用随机森林分类器对水稻种子图像进行分类,准确率为 90.54%,以上方法忽略了多类型特征带来的冗余性会造成识别准确率下降的问题。Liu 等^[13]采用水稻的颜色和几何特征,利用神经网络对 PCA 后的特征向量进行识别,平均识别率为 84.83%。

为了降低水稻种子的识别成本,选择从图像上来获取水稻种子的特征参数。针对单一特征的水稻识别方法带来的信息丢失问题,采用水稻种子图像的颜色、几何及纹理特征,多种特征使原始数据包含了更多有效信息。针对相似水稻种子品种识别效率低的问题,利用运行速度快、识别准确率高的 Bayes 模型对数据进行分类识别。采用 LDA(线性判别分析)算法对特征分析降维,去除原始数据中的无用信息和融合多特征数据的显著信息,解决了多特征带来的数据冗余问题,提高了识别模型性能。作为对比,试验将原始数据和 PCA、因子分析(FA)、LDA、局部线性嵌入(LLE)降维数据用于 Bayes、K-邻近(KNN)、SVM、多层感知机(MLP)分类模型中进行研究,通过图片增强技术丰富数据集的多样性来提高 LDA_Bayes 模型的适应能力,最后证明基于图像增强的 LDA_Bayes 模型能够优化传统的 Bayes 模型,提高水稻种子品种的识别效率,为水稻品种识别提供新方法。

1 材料与方法

1.1 试验材料

试验选取 4 类水稻种子(图 1)样本进行研究,品种为楚粳 7 号、马坝油粘、玉杨糯、玉针香,每类水稻种子选取成熟且颗粒饱满的样本 500 粒。选用佳能 EOS 60D 型数码相机(分辨率为 5 184 像素×3 456 像素)在水稻种子正上方进行垂直拍摄。图像处理和分析在中央处理器(CPU)为 Intel(R) Core(TM) i5-6200U、内存为 8 G 的计算机上进行。将水稻种子原始图像裁剪为分辨率为 1 616 像素×720 像素,

进行图像分割处理,使用高斯模糊对图像进行降噪。



图 1 水稻种子样本图像

Fig.1 Sample images of rice seeds

1.2 数据增强

为提高模型泛化能力,采用多种算法模拟不同测试环境对图像数据集进行增强,使训练模型能够学习到更多图片的像素信息。试验通过随机调节水

稻种子图像的亮度、对比度、色度和增强图像的锐化程度将水稻种子样本数量扩充到原来的 4 倍。增强样例如图 2 所示。



图 2 增强算法效果

Fig.2 Effects of enhanced algorithm

1.3 提取图像特征

为了使试验具有全面性,提取水稻种子图像的颜色通道(R、G、B)、正交投影面积、正交投影周长、长轴长、短轴长、二阶矩、熵、对比度、反分差矩阵共 11 个特征向量作为水稻种子品种识别的特征数据。

1.3.1 颜色特征 颜色特征具有良好的鲁棒性,是能体现水稻种子类别的重要特征参数。RGB 颜色空间是生活中常用的一种颜色空间模型,本研究通过计算水稻种子图像的 R、G、B 3 个分量的平均值作为颜色特征数据。

1.3.2 几何特征 本研究通过阈值算法对水稻种子图像进行分割,利用水稻种子图像的区域和边缘特性提取几何特征。

正交投影面积:计算分割后的水稻种子图像区域所包含的像素点的总和。

正交投影周长:通过 Sobel 算子计算图像梯度,与 Canny 算法相结合,提取水稻种子图像的边缘,计算边缘像素点的和。

长轴长、短轴长:通过外接矩阵框计算水稻种子 2 个端点的最长距离为长轴长,垂直于长轴的直线距离为短轴长。

1.3.3 纹理特征 纹理特征是一种用来描述图像表面性质的全局特征。试验通过提取水稻种子灰度

共生矩阵纹理(GLCM)^[14-15]的 4 个经典统计量:二阶矩(ASM)、熵(ENT)、对比度(CON)、反差分矩阵(IDM),将其作为纹理特征向量,计算方法分别如公式(1)、公式(2)、公式(3)和公式(4)所示。二阶矩反映图像的均匀性,值越大,说明图像灰度分布越均匀;熵是描述图像的信息量,值越大,说明图像的元素分布越分散;对比度用于衡量图像局部灰度变化程度,值越大,说明图像的边缘越锐利;反差分矩阵用于测量纹理的清晰程度,值越大,说明图像纹理越清晰和易于描述。

$$ASM = \sum_i \sum_j P(i, j)^2 \quad (1)$$

$$ENT = - \sum_i \sum_j P(i, j) \lg [P(i, j)] \quad (2)$$

$$CON = \sum_i \sum_j (i-j)^2 P(i, j) \quad (3)$$

$$IDM = \sum_i \sum_j \frac{P(i, j)}{1 + (i-j)^2} \quad (4)$$

式中, $P(i, j)$ 表示灰度图像中坐标为 (i, j) 的像素点的灰度值。

1.4 方法描述

1.4.1 LDA 算法 LDA 算法^[16]是一种有监督的机器学习,它的主要思想是把数据投影到低维空间,让属于一类的数据分布得更加紧密,使不同类的数据点更加分散,具有计算速度快的优点。该算法的流程如下:

Step1: 设数据集有 M 个类别: $D1, D2, \dots, Di$ 。
计算每类样本的类内散度矩阵 S_w :

$$S_w^i = \frac{1}{N} \sum_{k=1}^N (X_k - m^i)(X_k - m^i)^T \quad (5)$$

式中, S_w^i 表示第 i 类样本的类内散度矩阵, N 表示第 i 类样本的样本数, m^i 表示第 i 类样本的总体均值, X_k 表示第 i 类样本的第 k 个特征向量。

Step2: 计算各类样本之间的类间散度矩阵 S_b :

$$S_b^{ij} = \sum_{i=1}^M \sum_{j=1}^M (m^i - m^j)(m^i - m^j)^T \quad (6)$$

式中, S_b^{ij} 表示第 i 类样本与第 j 类样本之间的类间散度矩阵, m^j 表示第 j 类样本的总体均值。

Step3: 计算每个类别 LDA 转化矩阵 W_L 的矩阵值。提取矩阵值中最大的 d 个特征值和与其对应的特征向量, 得到投影矩阵:

$$W_L = S_w^{-1} S_b \quad (7)$$

式中, W_L 是 LDA 的转换矩阵, S_w^{-1} 表示该样本类内散度矩阵的逆矩阵, S_b 表示该样本的类间散度矩阵。

Step4: 每个样本可通过公式 (8) 转化为新的样本集, 则为降维后的数据集:

$$y = W_L^T x \quad (8)$$

式中, y 表示新的数据集, x 表示每个样本的特征值。

1.4.2 Bayes 分类器 Bayes 分类器^[17-18]的原理是已知各个类别在样本空间中出现的先验概率和某个类别中特征向量 X 出现的条件概率密度, 计算该类别在 X 出现的条件下, 每个样本属于各个类别的概率, 最后将样本分配到概率最大的那类数据集中。其计算流程如图 3 所示。

1.4.3 方法步骤 本研究提出基于 LDA 和 Bayes 相结合的模型 (LDA_Bayes) 对水稻种子的特征数据进行分类识别, 步骤如下:

Step1: 对水稻种子原始图像进行裁剪、分割、增强等预处理; Step2: 提取水稻种子图像的颜色特征、几何特征、纹理特征共 11 个特征向量; Step3: 采用 LDA 降维算法将水稻种子的 11 维特征向量降维到 2 维; Step4: 采用 Bayes 分类器对降维后 80% 的数据进行分类训练, 20% 的数据用于测试识别; Step5: 将试验模型与原始数据、增强数据和 15 类组合模型进行对比研究, 采用精确率、召回率、准确率、 $F1$ 作为评价指标。

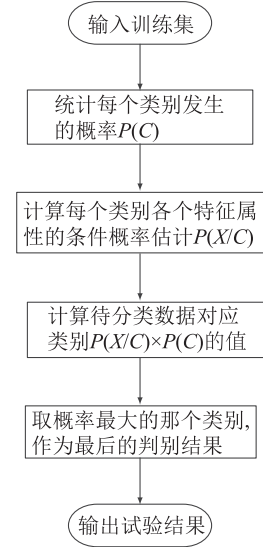


图 3 Bayes 分类流程

Fig.3 Flowchart of Bayes classification

2 结果与分析

2.1 原始数据集分类结果

表 1 为 11 个特征向量在 KNN、SVM、Bayes、MLP^[19-20] 分类器上的分类结果, 可以看出 Bayes 模型运行时间最短, 为 0.016 s, 测试准确率高达 97.3%, 说明 Bayes 分类器相较于其他 3 种分类器更适合试验数据分类。

2.2 降维数据分析与分类结果

2.2.1 数据降维分析可视化 为了验证 LDA 特征降维的有效性, 试验还利用 PCA、FA、LLE^[21-24] 降维方法对水稻种子图像的颜色、几何、纹理等 11 个特征向量进行分析降维, 将水稻种子的 11 维特征融合降到 2 维, 经 PCA 降维后数据的第 1 主成分贡献率为 98.41%, 第 2 主成分贡献率为 1.45%, 累积贡献率为 99.86%, 从图 4a 可以看出, 各样本之间重叠部分较大。经 LDA 降维后的数据第 1 判别因子的贡献为 69.86%, 第 2 判别因子的贡献率为 26.73%, 累积贡献率为 96.59%, 从图 4c 可以看出, 样本之间分布的距离较大, 重叠部分少。从图 4b、4d 可以看出, 用 FA、LLE 对数据降维后, 样本分布较为密集, 样本之间的间距小。

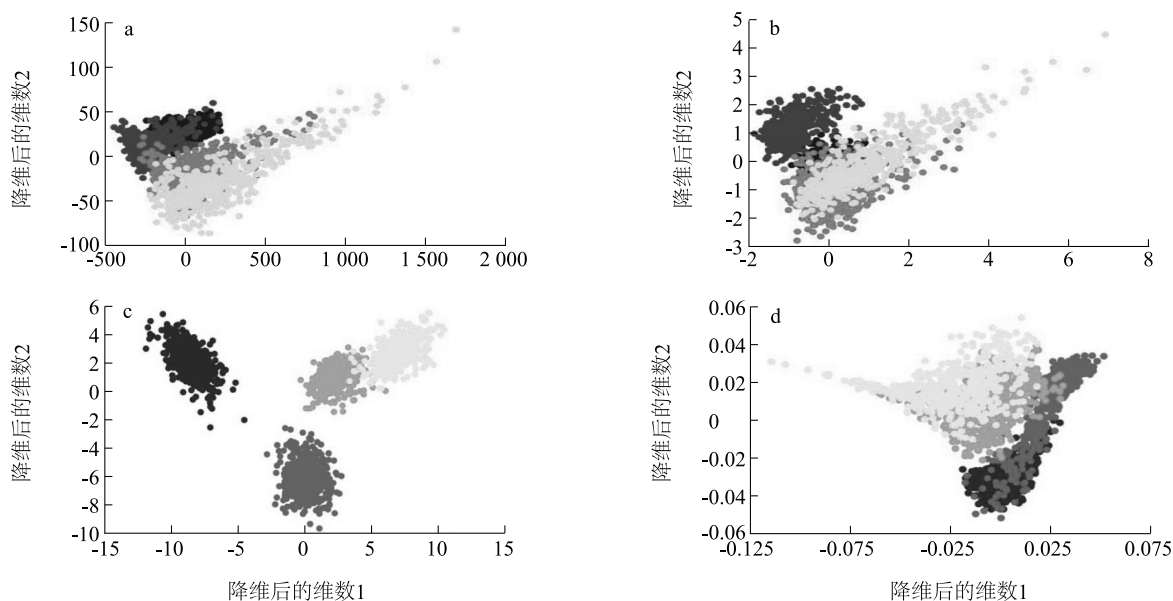
2.2.2 降维数据分类结果比较 为了验证 LDA_Bayes 模型的高效性, 试验利用降维数据进行 KNN、SVM、MLP 模型分类识别。

表 1 水稻种子原始数据分类结果

Table 1 Classification results of the original data of rice seeds

模型	时间 (s)	训练准确率 (%)	测试集			
			准确率(%)	精确率(%)	召回率(%)	<i>F1</i> (%)
KNN	0.088	93.2	88.8	88.7	88.8	88.7
SVM	0.321	77.6	77.5	85.7	77.5	72.6
Bayes	0.016	97.7	97.3	97.3	97.3	97.3
MLP	0.473	81.4	82.8	85.4	82.8	80.2

KNN、SVM、Bayes、MLP 为分类器。*F1*:模型精确率和召回率的调和平均数。



a:PCA 对水稻种子数据降维;b:FA 对水稻种子数据降维;c:LDA 对水稻种子数据降维;d:LLE 对水稻种子数据降维。

图 4 水稻种子特征数据降维结果

Fig.4 Dimensionality reduction results of feature data of rice seeds

表 2 为原始水稻种子特征数据经过 PCA 降维后结合 KNN、SVM、Bayes、MLP 分类器处理的试验结果,可以看出,PCA 降维是从特征的协方差角度进行考虑的,适合保持样本内在信息的数据降维,试验数据经过 PCA 降维后缩短了各模型的运行时间,但测试准确率并没有提高,说明 PCA 降维不适合处理本研究数据。

表 3 为原始水稻种子特征数据经过 FA 降维后结合 KNN、SVM、Bayes、MLP 分类器处理的试验结果,可以看出,与原始数据的运行时间和测试准确率相比,分类效率不仅没有提高,反而降低,所以该降维方法不适用于本研究数据降维。

表 4 为原始水稻种子特征数据经过 LDA 降维后结合 KNN、SVM、Bayes、MLP 分类器处理的试验结果,可以看出,与原始数据(表 1)相比,LDA_KNN

的测试准确率提高了 10.4 百分点,LDA_SVM 的测试准确率提高了 21.5 百分点,LDA_Bayes 的测试准确率提高了 2.3 百分点,LDA_MLP 的测试准确率提高了 16.6 百分点,说明 LDA 对试验数据降维时能够最大限度地区分不同类样本,在从高维空间降维到低维空间时可以最大限度地减少重要特征信息的丢失,提高归类能力。

表 5 为原始水稻种子特征数据经过 LLE 降维后结合 KNN、SVM、Bayes、MLP 分类器处理的试验结果,可以看出,LLE 适合分布均匀且密集的数据,适合流形数据的降维,试验数据经过 LLE 降维后测试准确率并没有提高,说明 LLE 降维方法不适合处理本研究数据。

综上所述,PCA、FA、LLE 3 种降维方法没有实现水稻种子有效信息的综合提取。利用 LDA 算法对

水稻种子特征数据降维后结合 4 种分类器处理的试验结果表明,该算法能够集中提取特征数据中的有效信息,摒弃无用信息,所以 LDA 算法降维比其他降维方法更适用于本研究中水稻种子的特征数据。试验数据经过 KNN、SVM、MLP、Bayes 分类器进行分类训练的识别结果表明, Bayes 能快速准确地对本研究中

的 4 类水稻种子进行分类识别。在原始数据和经降维处理的 16 种分类模型中, LDA_Bayes 模型在水稻种子分类识别中的运行时间为 0.013 s, 测试准确率为 99.6%, 相较于原始数据, 缩短了运行时间, 提高了测试准确率、精确率、召回率和 *F1* 等评价指标, 说明该模型具有高效性和可行性。

表 2 主成分分析(PCA)降维数据分类结果

Table 2 Classification results of PCA(Principal component analysis) dimensionality reduction data

模型	时间 (s)	训练准确率 (%)	测试集			
			准确率(%)	精确率(%)	召回率(%)	<i>F1</i> (%)
PCA_KNN	0.078	84.7	80.0	80.0	80.0	79.8
PCA_SVM	0.276	75.5	74.0	73.6	74.0	73.6
PCA_Bayes	0.016	79.1	79.8	81.3	79.8	79.2
PCA_MLP	0.277	79.6	78.9	78.7	78.9	78.7

F1 见表 1 注。

表 3 因子分析(FA)降维数据分类结果

Table 3 Classification results of FA(Factor analysis) dimensionality reduction data

模型	时间 (s)	训练准确率 (%)	测试集			
			准确率(%)	精确率(%)	召回率(%)	<i>F1</i> (%)
FA_KNN	0.090	78.0	69.5	70.1	69.5	68.8
FA_SVM	0.078	66.9	66.9	69.4	66.9	66.9
FA_Bayes	0.016	71.2	71.2	72.8	71.2	70.3
FA_MLP	1.139	75.5	75.3	75.8	75.3	74.9

F1 见表 1 注。

表 4 线性判别分析(LDA)降维数据分类结果

Table 4 Classification results of LDA(Linear discriminant analysis) dimensionality reduction data

模型	时间 (s)	训练准确率 (%)	测试集			
			准确率(%)	精确率(%)	召回率(%)	<i>F1</i> (%)
LDA_KNN	0.078	99.0	99.2	99.2	99.2	99.2
LDA_SVM	0.078	98.6	99.0	98.9	99.0	99.0
LDA_Bayes	0.013	98.6	99.6	99.6	99.6	99.6
LDA_MLP	1.248	98.7	99.4	99.4	99.4	99.4

F1 见表 1 注。

表 5 局部线性嵌入(LLE)降维数据分类结果

Table 5 Classification results of LLE(Locally linear embedding) dimensionality reduction data

模型	时间 (s)	训练准确率 (%)	测试集			
			准确率(%)	精确率(%)	召回率(%)	<i>F1</i> (%)
LLE_KNN	0.075	85.4	81.2	81.2	81.2	81.0
LLE_SVM	0.016	67.4	66.3	74.4	66.3	57.1
LLE_Bayes	0.016	78.3	77.7	77.5	77.7	77.3
LLE_MLP	0.984	80.9	78.9	78.7	78.9	78.5

F1 见表 1 注。

2.3 进行数据增强后的试验

表 6 为水稻种子特征数据增强前后结合 LDA_Bayes 模型的试验结果,可看出增强后数据集的运行时间为 0.019 s,训练准确率为 99.4%,相较于原始数据集的训练准确率有所提高,测试准确率为

99.4%,相较于原始数据集的测试准确率轻微下降,其原因为玉杨糯和玉针香 2 类样本特征过于相似,通过对数据的扩充使数据量已经使模型的数据信息达到饱和,因此通过增加训练数据集的方法难以提高识别准确率。

表 6 数据集增强前后结合 LDA_Bayes 模型的对比分类结果

Table 6 Comparison results before and after data set enhancement combined with LDA_Bayes model

数据集	时间 (s)	训练准确率 (%)	测试集			
			准确率(%)	精确率(%)	召回率(%)	F1(%)
原始数据集	0.013	98.6	99.6	99.6	99.6	99.6
增强后的数据集	0.019	99.4	99.4	99.4	99.4	99.4

F1 见表 1 注。

3 讨论

图像特征提取是图像分类识别的重要步骤,只进行单一的图像特征提取会丢失部分有效特征,如熊利荣等^[25]只将水稻种子图像的几何特征结合 Bayes 分类器进行识别,平均识别准确率为 81.4%,而本研究采取了易于反映图像显著信息的颜色特征、几何特征以及纹理特征,通过 Bayes 对 11 个特征向量进行分类识别,测试识别准确率为 97.3%。

特征数据过多会造成数据的冗余性,产生干扰信息,降低分类准确率。如黄星奕等^[11]提取水稻种子的颜色特征和形状特征等 13 个特征参数,利用贝叶斯分类器对原始数据进行分类,识别正确率为 88.3%。通过降维可以解决数据冗余性问题、去除干扰信息和提高识别准确率,因此降维方法的选择是提高分类模型性能的关键。相较于无监督学习的 PCA 算法,LDA 属于监督学习,在降维过程中可以使用类别的先验知识经验选择分类性能最好的投影方向;FA 利用相关性对特征进行降维,能够保存原始数据所存在的重要信息,但是不能像 LDA 一样摒弃一些对分类具有干扰的信息;相较于 LLE 只能适用于流形数据的局限性,采用基于线性数据处理的 LDA 算法更适用于本研究数据。因此本试验选取 LDA 算法对数据进行分析,将数据从高维空间最大区分化地降低到低维空间,有效融合了原始数据中的显著特征,提高了识别效率。

分类模型的选择是提高识别效率的关键,常用的分类模型有 KNN、SVM、MLP。KNN 是通过计算数据间的距离进行分类,对类别分类状况敏感,分类

识别结果易受 K 值影响;SVM 更适用于二分类算法,对大规模的数据集进行训练时计算复杂度高,效率不高;MLP 需要多次迭代,耗时长,会降低识别效率。因此 Bayes 运行时间短和识别效率高的优点更适合本研究数据分类。在特定场景和实际应用中,往往很难收集满足各种条件的数据用于训练和测试,因此本研究通过调节亮度、对比度、色度和锐化程度来丰富数据库种类以解决测试环境过于单一的问题,提高实际测试的鲁棒性。试验结果表明,基于图像增强的 LDA_Bayes 模型具有更强的泛化能力和适应性,能进一步提高水稻种子识别率和运行效率,可以为后期的在线实时识别水稻种子系统提供基础。

本研究的对象是楚粳 7 号、马坝油粘、玉杨糯和玉针香 4 类水稻种子,为提高试验模型的可靠性,将在后期的研究中加入其他品种的水稻种子。本研究下一步将对水稻种子的其他特征参数和特征降维方法作进一步研究,实现易于推广且效率高的在线实时识别水稻种子的系统。

4 结论

试验提取水稻种子图像的颜色特征、几何特征、纹理特征等共 11 个特征向量并采用 PCA、FA、LDA、LLE 4 种算法进行特征数据的分析降维,选择 KNN、MLP、SVM、Bayes 分类器进行数据增强前后分类识别的对比验证,结果表明:

(1) LDA 降维算法可以有效地对水稻种子图像所提取的 11 个特征参数进行降维,降维后的结果能够从多种特征中去除不同水稻种子之间相似的无

用特征,准确提取和融合显著特征。相较于传统的 Bayes 模型,LDA_Bayes 模型在测试准确率和运行速度上都有所提升。

(2)通过图像增强技术模拟了实际生活中的多种检测环境,增强了图像样本的鲁棒性,结果表明,基于图像增强的 LDA_Bayes 模型的训练准确率为 99.4%,测试准确率为 99.4%,运行时间为 0.019 s,说明数据增强能够有效增强模型的泛化能力和适应性,使该模型能在实际应用中高效识别种子,满足普遍推广和在线识别的要求。

参考文献:

- [1] MEBATSION H K, PALIWAL J, JAYAS D S. Automatic classification of non-touching cereal grains in digital images using limited morphological and color features[J]. *Computer&Electronics in Agriculture*, 2013, 90(90): 99-105.
- [2] CHOUDHARY R, PALIWAL J. Classification of cereal grains using wavelet, morphological, colour, and textural features of non-touching kernel images[J]. *Biosystems Engineering*, 2008, 99(3): 330-337.
- [3] 徐 赛,周志艳,罗锡文. 常规稻与杂交稻谷的仿生电子鼻分类识别[J]. *农业工程学报*, 2014, 30(9): 133-139.
- [4] 李 辉,王阳恩,刘 庆,等. 分段激光诱导击穿光谱的水稻种子识别[J]. *激光杂志*, 2017, 38(2): 8-12.
- [5] HAI V. Spatial and spectral features utilization on a hyperspectral imaging system for rice seed varietal purity inspection[C]//MACIEJ P, AGNIESZKA T. 2016 IEEE RIVF International Conference on Computing & Communication Technologies. Hanoi: IEEE, 2016(10): 169-174.
- [6] FABIYI S, HAI V, TACHTATZIS C, et al. Varietal classification of rice seeds using RGB and hyperspectral images[J]. *IEEE Access*, 2020, 8: 22493-22505.
- [7] 邓小琴,朱启兵,黄 敏. 融合光谱、纹理及形态特征的水稻种子品种高光谱图像单粒鉴别[J]. *激光与光电子学进展*, 2015, 52(2): 128-134.
- [8] 李 冬. 基于图像的稻花香水稻种子鉴别方法的研究[D]. 大庆: 黑龙江八一农垦大学, 2018.
- [9] 郑 宇. 基于机器视觉的稻谷种子特征提取与品种识别方法研究[D]. 武汉: 华中农业大学, 2009.
- [10] 李秀英,马 旭,李泽华,等. 基于特征融合和 SVM 的稻谷品种识别[J]. *中国农机化学报*, 2019, 40(7): 97-102.
- [11] 黄星奕,李 剑,姜 松. 基于计算机视觉的稻谷品种识别技术的研究[J]. *江苏大学学报(自然科学版)*, 2004, 25(2): 102-104.
- [12] HONG P, HAI T, LAN L. Comparative study on vision based Rice seed varieties identification[C]//NGUYEN L M, NGUYEN T T. 2015 Seventh International Conference on Knowledge and Systems Engineering. Ho Chi Minh City: IEEE, 2015, 11(7): 377-382.
- [13] LIU Z Y, CHENG F, YING Y B, et al. Identification of rice seed varieties using neural network[J]. *Journal of Zhejiang University Science B*, 2005, 6(11): 53-58.
- [14] BARALDI A, PARMIGIANI F. An investigation of the textural characteristics associated with GLCM matrix statistical parameters[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 1995, 33(2): 293-304.
- [15] HARALICK R M, SHANMNGAM K, DINSTEN I. Texture feature for image classification[J]. *IEEE Transactions on Systems, Man and Cybernetics*, 1973, 3(6): 768-780.
- [16] 章万锋. 基于 PCA 与 LDA 的说话人识别研究[D]. 杭州: 浙江大学, 2004.
- [17] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [18] NIKITA E, NIKITAS P. Sex estimation: a comparison of techniques based on binary logistic, probit and cumulative probit regression, linear and quadratic discriminant analysis, neural networks, and naïve Bayes classification using ordinal variables[J]. *International Journal of Legal Medicine*, 2020, 134(3): 1213-1225.
- [19] FAYYAZI S, ABBASPOURFARD M H, ROHANI A, et al. Identification and classification of three Iranian rice varieties in mixed bulks using image processing and MLP neural network[J]. *International Journal of Food Engineering*, 2017, 13(5).
- [20] ZANATY E A. Support vector machines (SVMs) versus Multilayer perception (MLP) in data classification[J]. *Egyptian Informatics Journal*, 2012, 13(3): 177-183.
- [21] 彭彦昆,赵 芳,李 龙,等. 利用近红外光谱与 PCA-SVM 识别热损伤番茄种子[J]. *农业工程学报*, 2018, 34(5): 159-165.
- [22] 梁晓莹,田光杰. 基于混合特征 LLE 融合与 SVM 的质量异常模式识别[J]. *组合机床与自动化加工技术*, 2020(3): 55-59, 64.
- [23] 刘开南,冯新扬,邵 超. 一种面向图像分类的流形学习降维算法[J]. *计算机应用与软件*, 2019, 36(8): 210-213, 229.
- [24] 高惠璇. 应用多元统计分析[M]. 北京: 北京大学出版社, 2005.
- [25] 熊利荣,郑 宇. 基于形态学的稻谷种子品种识别[J]. *粮油加工*, 2010(6): 45-48.

(责任编辑:陈海霞)