基于高光谱与极限学习机的甜菜品种识别研究

**摘 要：**本文结合高光谱技术及极限学习机（Extreme Learning Machine，ELM）算法实现了对甜菜品种的识别，并研究了卷积平滑、一阶导数、卷积平滑与一阶导数、对数变换、对数变换与一阶导数、标准正态变换、标准正态变换与一阶导数7种预处理方法对ELM模型识别性能的影响。为简化输入变量，采用标准差峰值法对不同预处理下的光谱数据提取了特征波段。实验结果表明，在不同预处理方法下，ELM模型对甜菜品种的识别率均在80%以上，表现出良好的学习性能和泛化性能，且发现不同预处理方法与一阶导数结合可以有效提高光谱数据的信噪比，增强光谱模型的准确性和稳定性；以特征波段建立的ELM模型对甜菜品种的识别精度比全波段建立的模型在总体上有所提高，表明以标准差峰值法提取的特征波段是有效的；基于卷积平滑与一阶导数预处理后的光谱数据在全波段及特征波段下所建立的ELM模型均取得了最优的识别效果，其训练集的识别率均达到了98.13%，预测集的识别率分别为93.94%、95.45%。

**关键词：**高光谱；极限学习机；甜菜品种

**中图分类号：**  **文献标识码：**A  **文章编号：**

引言

高光谱技术是20世纪80年代兴起的一种快速、无损的综合性探测技术，其具有光谱分辨率高、波段多且连续的特点,可以反映出目标物体相邻波段间的细微光谱差异[1]，克服了多光谱技术宽波段、低分辨率的局限性，为研究地物之间的细微特征提供了有效的技术支撑。

近年来，国内外学者利用高光谱技术结合机器学习算法在对植物品种的识别方面进行了大量的研究工作，并取得了显著成果。张初等[2]利用高光谱技术对黑豆品种进行了鉴别研究，发现以小波变换提取的光谱特征信息建立的极限学习机（ELM）模型的识别效果最好，其建模集和预测集识别正确率均达到100%；马惠玲等[3]基于高光谱技术实现了对苹果品种的快速鉴别，建立的K 近邻（KNN）模型的识别精度达到了100%；Huang等[4]采用高光谱技术对不同年份的玉米种子进行识别，运用增量支持向量数据描述（ISVDD）算法对最小二乘支持向量机（LSSVM）进行了优化，其识别精度达到了94.4%。此外，有的学者利用高光谱技术对西瓜种子[5]、大白菜种子[6]、红豆[7]、稻米[8]、荔枝[9]等植物的品种进行了识别研究，并取得了较好的效果，而对甜菜品种建立高光谱识别模型的研究还未见报道。

甜菜是我国主要的制糖原料之一，也是黑龙江、内蒙古及新疆等地重要的经济作物。甜菜品种与甜菜栽培、制糖工业有着极为密切的联系[10]，正确选择适合本地区种植的甜菜品种以及甜菜品种之间的合理搭配种植能够有效提高甜菜的产量及含糖率[11]，因此对甜菜品种进行快速识别对甜菜品种的选育及田间的生产管理具有重要的指导作用。本文以3个甜菜品种为试验对象，采用高光谱技术和极限学习机（ELM）算法对3个甜菜品种建立快速识别模型，并研究不同的高光谱预处理方法对ELM模型识别精度的影响，从中筛选出最优模型，从而为甜菜的生产管理提供理论依据。

1 材料与方法

1.1 试验材料

试验选用的甜菜品种为SD21816、KWS1676、KWS9147，试验采用大田小区方式进行，其中SD21816于2013年种植于内蒙古呼和浩特土左旗北什轴乡境内甜菜种植基地，共种植9个小区；KWS1676于2014年种植于内蒙古赤峰市松山区太平地镇甜菜规模种植区，共种植28个小区；KWS9147于2015年种植于内蒙古农业大学试验田，共种植18个小区。试验中于6月中下旬至9月中下旬多次进入田间采集甜菜冠层光谱数据，3个甜菜品种共采集了400个样本数据。在本研究中，将样本数据以2:1的比例随机划分为训练集（268个）和预测集（132个），并以赋值法来确定甜菜品种，将品种SD21816、KWS1676、KWS9147分别赋值为1、2、3，其具体划分情况如表1所示。

表1 甜菜样本划分情况

|  |  |  |  |
| --- | --- | --- | --- |
| 甜菜品种 | 赋值 | 训练集 | 预测集 |
| SD21816 | 1 | 72 | 35 |
| KWS1676 | 2 | 112 | 56 |
| KWS9147 | 3 | 84 | 41 |

1.2 冠层光谱数据采集

本试验采用美国ASD 公司生产的野外便携式 ASD Qualityspec 高光谱仪对甜菜冠层光谱进行采集。光谱仪波段范围为 350-1830 nm，其中，350-1000 nm采样间隔为 1.4 nm，光谱分辨率为 3 nm；1000-1830 nm 光谱采样间隔为 2 nm，光谱分辨率为 10 nm。采集时间选择在天气晴朗、无风、无云的北京时间10:00-14:00 之间进行。采集光谱时光纤探头垂直向下，距甜菜冠层的垂直高度依据样品冠层大小及探头视场角(25°)确定，使样品冠层恰能位于探头视场范围内。在每个小区选择具有代表性的样本点并连续采集4次光谱，将平均值作为该样品光谱，且在每个试验小区采集光谱前进行参考白板校正。

1.3 光谱数据预处理

光谱仪所采集的光谱数据中，除了包含与样品自身有关的信息外，还包含了因多种因素产生的杂散光、土壤背景等噪声信号。这些噪音信号会影响光谱信息的真实性，从而影响所建模型的准确性和稳定性，因此对光谱数据进行预处理是非常必要的。然而，并不是每一种预处理方法都能达到很好的效果。

本研究利用卷积平滑、一阶导数、卷积平滑与一阶导数、对数变换、对数变换与一阶导数、标准正态变换、标准正态变换与一阶导数7种预处理方法对高光谱数据进行预处理，并对7种预处理方法进行对比分析。

1.4 特征波段选择

由于高光谱的波段数多、数据量大，相邻波段间的相关性强，从而影响数据的处理速度，给研究带来了难度，也在一定程度上限制了高光谱技术的广泛应用[12-13]。因此，在保证模型精度的前提下，对高光谱数据进行降维是非常必要的。国内外学者通常利用光谱特征提取和特征选择的方法对高光谱数据进行降维[14-16]。

本研究采用标准差峰值法来选取甜菜冠层光谱的特征波段，其基本思想是：所有样本的光谱数据在各个波段下的标准差越大，表明在此波段下，甜菜不同品种之间的光谱差异就越显著，区分甜菜品种的能力就越强，而标准差曲线上的峰值是相邻波段间标准差的最大值，其蕴含甜菜不同品种之间的光谱信息也是最大的。

其筛选过程如下：

（1）计算所有甜菜样本在各个波段下的标准差，其计算公式如下：

 (1)

其中，为第个波段下的标准差，为样本个数，为第个波段下的光谱平均值。

（2）筛选出标准差曲线上的各个峰值，计算出所有峰值的平均值；

（3）将标准差曲线上高于的峰值所对应的波段作为特征波段。

1.5 识别模型

极限学习机（Extreme Learning Machine，ELM）[17]是黄广斌教授在单隐含层前馈神经网络（SLFNs）的基础上提出的一种新型的机器学习算法，由输入层、隐含层和输出层组成。该算法通过随机产生输入层与隐含层之间的输入权重w和隐含藏的偏置b，并且在训练过程中无需调整，只需设置隐含层神经元的激励函数g(x)及个数L，就能唯一确定隐含层与输出层间的输出权重β，从而获得唯一的最优解。ELM算法具有学习速度快、泛化性能好等优点，克服了传统神经网络算法在训练过程中不断调整网络参数的弊端，改善了传统神经网络存在的训练速度慢、陷入局部而非全局最优、学习率选择敏感等问题[18]。

在本研究中，ELM隐含藏的激励函数选择sigmoid 函数，隐含层神经元的个数设置为60。

1.6 光谱数据分析软件

在本研究中，对光谱数据的各种预处理、特征波段的选择以及识别模型的建立均在MATLAB R2014a软件中完成。

2 结果与分析

2.1 甜菜冠层原始光谱响应分析



图1 甜菜冠层原始反射光谱曲线

图1为甜菜全部样本的原始反射光谱。由图1可知，3个品种的反射光谱曲线走势基本相同，并没有明显差异，而在反射率的高低上有所不同，因此对不同甜菜品种的识别还需要进一步的建模分析；同时可以看出，在1351-1450nm波段处及光谱末端有明显噪声，考虑到原始光谱的首端也会含有大量的随机噪声[19]，因此在本研究中只选取400-1350nm之间的反射光谱数据用于后续预处理及建模分析。

2.2 基于全波段的甜菜品种识别

利用上述7种预处理方法对400-1350nm之间的原始光谱数据进行预处理，而后利用原始及7种预处理后的光谱数据对甜菜品种建立ELM识别模型，其识别结果如表2所示。

表2 基于全波段的ELM模型识别结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 预处理方法 | 训练集 | | 预测集 | |
| 正确  识别数 | 识别率/% | 正确  识别数 | 识别率/% |
| 原始 | 258 | 96.27 | 117 | 88.64 |
| 卷积平滑 | 259 | 96.64 | 117 | 88.64 |
| 一阶导数 | 260 | 97.01 | 121 | 91.67 |
| 卷积平滑+一阶导数 | 263 | 98.13 | 124 | 93.94 |
| 对数变换 | 250 | 93.28 | 108 | 81.82 |
| 对数变换+一阶导数 | 263 | 98.13 | 123 | 93.18 |
| 标准正态变换 | 244 | 91.04 | 109 | 82.58 |
| 标准正态变换+一阶导数 | 260 | 97.01 | 122 | 92.42 |

由表2可知，基于全波段的不同预处理方法下所建的ELM模型对甜菜品种的识别率均达到了80%以上。其中，经卷积平滑与一阶导数预处理后的光谱数据所建立的ELM模型识别效果最好，其训练集的识别率为98.13%，预测集的识别率为93.94%。经卷积平滑的识别效果与原始光谱的识别效果相当，训练集的识别率分别为96.64%、96.27%，预测集的识别率均为88.64%。经对数变换、标准正态变换的识别效果较差，训练集与预测集的识别率均低于原始光谱的识别率。其余预处理方法所建的ELM模型的识别效果均优于原始光谱的识别效果。

从预处理方法的结合情况来看，不同预处理方法（卷积平滑、对数变换、标准正态变换）与一阶导数结合对原始光谱预处理后所建立的ELM模型对甜菜品种的识别效果均优于单种预处理方法的识别效果。这表明，在全波段下，不同预处理方法与一阶导数结合可以有效提高光谱数据的信噪比，增强光谱模型的准确性和稳定性。

2.3 光谱特征波段的选取

本研究截取的400-1350nm之间的光谱数据共951个光谱波段，为降低光谱信息的冗余度，本研究利用标准差峰值法对不同预处理下的光谱数据进行特征波段选取，其选择结果如表3所示。

表3 基于标准差峰值法选择的特征波段个数

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 预处理方法 | 原始 | 卷积平滑 | 一阶导数 | 卷积平滑+一阶导数 | 对数变换 | 对数变换+一阶导数 | 标准正态变换 | 标准正态变换+一阶导数 |
| 特征波段个数 | 87 | 35 | 83 | 51 | 29 | 106 | 61 | 98 |

由表3可知，在不同预处理方法下所选择的特征波段个数均不相同，基于对数变换后的光谱数据选择的特征波段个数最少，为29个；基于对数变换与一阶导数处理的光谱数据选择的特征波段个数最多，为106个。相比于原始的951个光谱波段，经特征波段的选取后，光谱信息的维度被大大压缩。

2.4 基于光谱特征波段的ELM模型识别结果

利用不同预处理下所筛选的特征波段分别建立ELM模型，其对甜菜品种的识别结果如表4所示。

由表4可知，相比于全波段建立的ELM模型，除一阶导数外，其余预处理方法在特征波段下建立的ELM模型在训练集和预测集的识别率均有所提高。其中，基于卷积平滑与一阶导数结合的特征波段所建立的ELM模型的识别效果最优，训练集的识别率达到98.13%，预测集的识别率达到95.45%；基于一阶导数的特征波段所建立的ELM模型的识别率有所下降，但也达到了理想的识别精度，训练集的识别率为95.90%，预测集的识别率为89.39%。这表明，应用标准差峰值法所选择的特征波段基本涵盖了甜菜品种判别的有效信息，大大降低了光谱信息的变量数目，提高了数据的处理效率。

表4 基于特征波段的ELM模型识别结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 预处理方法 | 训练集 | | 预测集 | |
| 正确  识别数 | 识别率/% | 正确  识别数 | 识别率/% |
| 原始 | 264 | 98.51 | 126 | 95.45 |
| 卷积平滑 | 263 | 98.13 | 121 | 91.67 |
| 一阶导数 | 257 | 95.90 | 118 | 89.39 |
| 卷积平滑+一阶导数 | 263 | 98.13 | 126 | 95.45 |
| 对数变换 | 259 | 96.64 | 117 | 88.64 |
| 对数变换+一阶导数 | 262 | 97.76 | 125 | 94.70 |
| 标准正态变换 | 260 | 97.01 | 122 | 92.42 |
| 标准正态变换+一阶导数 | 261 | 97.39 | 124 | 93.94 |

从预处理方法的结合来看，基于不同预处理方法（卷积平滑、对数变换、标准正态变换）与一阶导数结合的特征波段建立的ELM模型识别效果仍然优于单种预处理方法下的识别效果，这与全波段下所得出的结论一致。

3 结论与讨论

本研究利用高光谱技术及ELM算法对甜菜的3个品种（SD21816、KWS1676、KWS9147）进行了识别研究，并对不同预处理方法下（卷积平滑、一阶导数、卷积平滑与一阶导数、对数变换、对数变换与一阶导数、标准正态变换、标准正态变换与一阶导数）ELM模型的识别性能进行了对比，现得出以下结论：

（1）在不同预处理方法下，ELM模型均取得了理想的识别精度，训练集和预测集的识别率均在80%以上，表明ELM算法在对甜菜品种的识别过程中表现出良好的学习性能和泛化性能；

（2）与全波段建立的ELM模型相比，利用特征波段建立的ELM模型对甜菜品种的识别精度在总体上有所提高，这表明应用标准差峰值法所提取的特征波段基本涵盖了甜菜品种判别的有效信息，从而精简了光谱变量，提高了信息的处理效率；

（3）在不同预处理方法下，基于卷积平滑与一阶导数预处理后的光谱数据在全波段及特征波段下所建立的ELM模型均取得了最优的识别效果，其训练集的识别率均为98.13%，预测集的识别率分别为93.94%、95.45%；

（4）从预处理方法的结合情况来看，不同预处理方法与一阶导数结合可以有效提高光谱数据的信噪比，增强光谱模型的准确性和稳定性。

本研究表明，利用高光谱技术对甜菜品种进行识别是可行的，但本研究仅对甜菜的3个品种建立了ELM识别模型，在今后的研究中应对更大范围的甜菜品种建立更为稳健的判别模型，从而为甜菜田间的生产管理提供更为有力的理论支撑。

参考文献

1. 浦瑞良, 宫鹏. 高光谱遥感及其应用[M]. 北京: 高等 教育出版社, 2000.
2. 张初, 刘飞, 章海亮, 等. 近地高光谱成像技术对黑豆品种无损鉴别[J]. 光谱学与光谱分析, 2014, 34(3): 746-750.
3. 马惠玲, 王若琳, 蔡骋, 等. 基于高光谱成像的苹果品种快速鉴别[J]. 农业机械学报, 2017, 48(4): 305-312.
4. Huang M, Tang J Y, Yang B, et al. Classification of maize seeds of different years based on hyperspectral imaging and model updating[J]. Computers and Electronics in Agriculture, 2016, 122: 139-145.
5. 张初, 刘飞, 孔汶汶, 等. 利用近红外高光谱图像技术快速鉴别西瓜种子品种[J]. 农业工程学报, 2013, 29(20): 270-277.
6. 程术希, 孔汶汶, 张初, 等. 高光谱与机器学习相结合的大白菜种子品种鉴别研究[J]. 光谱学与光谱分析, 2014, 34(9): 2519-2522.
7. 孙俊, 路心资, 张晓东, 等. 基于高光谱图像的红豆品种GA-PNN神经网络鉴别[J]. 农业机械学报, 2016, 47(6): 215-221.
8. WANG L, LIU D, PU H, et al. Use of hyperspectral imaging to discriminate the variety and quality of rice[J]. Food Analytical Methods, 2014, 8(2): 515-523.
9. LIU D, WANG L, SUN D, et al. Lychee variety discrimination by hyperspectral imaging coupled with multivariate classification[J]. Food Analytical Methods, 2014, 7(9): 1848-1857.
10. 沈玉和. 我国的甜菜品种与类型[J]. 甜菜糖业, 1990, (5): 52-54
11. 王燕飞, 张立明, 杨洪泽, 等. 新疆甜菜品种的合理配置[J].新疆农业科学, 2003, 40(S2): 69-71.
12. 刘秀英, 林辉, 熊建利, 等. 森林树种高光谱波段的选择[J]. 遥感信息, 2005, (4): 41-44+64.
13. 王旭红, 肖平, 郭建明. 高光谱数据降维技术研究[J]. 水土保持通报, 2006, 26(6): 89-91.
14. 邱琳, 林辉, 臧卓, 等. 基于均值置信区间带的湿地植被高光谱特征波段选择[J]. 中南林业科技大学学报, 2013, 33(1): 41-45.
15. Hsu P H. Feature extraction of hyperspectral images using wavelet and matching pursuit[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2007, 62(2):78-92.
16. Du Q, He Y. Similarity-Based Unsupervised Band Selection for Hyperspectral Image Analysis[J]. IEEE Geoscience and Remote Sensing Letters, 2008, 5(4): 564-568.
17. Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications[J]. Neurocomputing, 2006, 70(1/2/3): 489-501.
18. 黄锋华, 张淑娟, 杨一, 等. 油桃外部缺陷的高光谱成像检测[J]. 农业机械学报, 2015, 46(11): 252-259.
19. 赵旭婷, 张淑娟, 刘蒋龙, 等. 高光谱技术结合CARS-ELM的油桃品种判别研究[J]. 现代食品科技, 2017, 33(10): 281-287.