

孙俊,金夏明,毛罕平,等. 基于有监督特征提取的生菜叶片农药残留浓度高光谱鉴别研究[J]. 江苏农业科学,2014,42(5):227-229.

基于有监督特征提取的生菜叶片 农药残留浓度高光谱鉴别研究

孙俊¹, 金夏明¹, 毛罕平², 刘泉¹, 武小红¹, 方敏¹, 宗祥¹

(1. 江苏大学电气信息工程学院, 江苏镇江 212013; 2. 江苏大学江苏省现代农业装备与技术重点实验室, 江苏镇江 212013)

摘要:为了保证人们对蔬菜的安全食用,研究了蔬菜叶片农药残留的无损检测方法。标准营养液无土栽培生菜样本,在成熟期按 4 种不同浓度,分别为 1.250、0.830、0.600、0.375 mL/L,将氰戊菊酯农药雾状均匀喷洒至生菜叶片上,8 h 后采集生菜叶片高光谱数据。采用标准归一化(SNV)算法对原始光谱进行预处理,分别利用基于非监督特征提取方法主成分分析(PCA)、局部保留投影(LPP)与基于监督特征提取方法线性判别分析(LDA)、局部保留投影(SLPP)对降噪后的光谱数据进行特征提取,统一选用支持向量机(SVM)作为分类器。利用相同的训练样本与测试样本进行分类试验,对生菜叶片农药残留浓度分类鉴别的结果为,PCA-SVM 分类正确率为 82.14%,LPP-SVM 分类正确率为 85.71%,LDA-SVM 分类正确率为 89.29%,SLPP-SVM 分类正确率达到 92.86%。结果表明,与非监督特征提取算法相比,监督特征提取算法由于充分利用了样本的类别特性,使得分类器对降维后的数据更加敏感,分类精度更高,其中 SLPP-SVM 的分类效果最好。

关键词:生菜;监督;非监督;农药残留;高光谱

中图分类号: S126;S481⁺.8 **文献标志码:** A **文章编号:** 1002-1302(2014)05-0227-03

生菜含有丰富的蛋白质、维生素及矿物质,深受人们的喜爱。蔬菜的健康安全食用,一直是食品专家关注的问题。目前,蔬菜农残中毒事件屡次发生,如何快速检测蔬菜农药残留超标是当前蔬菜生产中食品安全专家亟待解决的问题之一。

目前,虽然对农药残留的检测已有较成熟的方法,如化学速测法、免疫分析法、仪器分析法[气相色谱法(GC)、液相色谱法(HPLC)等],但这些方法都存在样品前处理过程繁琐、试剂消耗大、耗时长等不足,无法实现快速与绿色检测的需要,不利于推广应用。基于光谱的无损检测技术具有快速、非破坏的特点,已成为当前国内外农作物信息检测的研究热点。关于光谱

技术的农残检测已有报道。陈蕊等利用在 600~1 100 nm 波段范围内可见-近红外反射光谱分析技术,对常见的高残留农药在绿色植物活体上的无损检测进行了研究^[1];Sanchez 等利用近红外反射光谱(NIRS)与偏最小二乘判别分析测定辣椒中的农药残留,无农药正确率与有农药判断正确率为 75% 和 82%,表明 NIRS 可以作为判断作物是否含有农药残留的初步筛选方法^[2]。

高光谱数据具有光谱分辨率高、相邻波段相关性强、信息冗余度大的特点,如何有效对高光谱数据进行降维一直是研究的重要方向。PCA、LPP、LDA、SLPP 作为 4 种经典的特征提取算法已经成功应用在人脸识别领域,其中 PCA 和 LPP 算法没有考虑到数据的类别信息,属于无监督特征提取算法^[3]。在实际生活或者试验中,一些样本的类别信息是已知的,在进行分类建模时,充分利用这些类别信息能有效提高分类模型的精度,LDA 和 SLPP 则利用了这一特点,属于有监督特征提取算法。本研究尝试将这 4 种特征提取算法引入到高光谱数据的降维中来,从有无监督的角度出发,分别利用这 4 种特征提取算法对高光谱数据进行特征提取,并利用 SVM 进

收稿日期:2013-08-13

基金项目:国家自然科学基金(编号:31101082,61075036);江苏高校优势学科建设工程(编号:苏政办发[2011]6号);国家级大学生创新创业训练计划(编号:201310299011);江苏省高等学校大学生创新创业训练计划(编号:201310299011Z)。

作者简介:孙俊(1978—),男,江苏泰兴人,博士,副教授,研究方向为计算机技术在农业工程中的应用。Email: sun2000jun@ujs.edu.cn。

试验结果均质压力为 25 MPa,蛋白质含量为 1.1%,pH 值为 3.65 组合条件最佳,在该条件下褐色乳酸菌饮料体系稳定,风味浓郁。

参考文献:

- [1] 陈健凯,陈健旋,林洵,等. 养乐多饮料中影响干酪乳杆菌代田株活菌数因素的研究[J]. 中国酿造,2008(22):34-36.
- [2] 徐致远,吴艳,周凌华,等. 影响褐色益生菌乳饮料颜色的因素[J]. 食品与发酵工业,2010,36(1):180-183.
- [3] 徐致远,吴艳,郭本恒,等. 一种褐色益生菌乳饮料的研制[J].

食品工业科技,2010,31(8):242-244.

- [4] 苏米亚,莫蓓红. 超高温牛奶产品货架期内的稳定性分析方法[J]. 乳业科学与技术,2006,29(5):221-223,226.
- [5] 郭翔,徐致远,艾连中,等. 褐色益生菌乳饮料稳定性研究及工艺条件优化[J]. 食品科学,2009,30(22):226-230.
- [6] 董涛,张国松,周玉玲. 养乐多风味乳酸菌饮料生产工艺的探讨[J]. 中国乳品工业,2005,33(4):40-41.
- [7] 姚晶,孟祥晨. 羧甲基纤维素钠及酸性乳饮料加工工艺对其稳定性及粒径分布的影响[J]. 中国乳品工业,2008,36(7):40-43.

行分类试验与比较。

1 试验与数据采集

1.1 样本种植与农药喷洒

试验品种选为意大利全年耐抽薹生菜,采用珍珠岩袋培无土栽培技术进行生菜样本培育,并利用营养液自动灌溉系统进行标准营养液灌溉。试验地点在江苏大学现代农业装备与技术省部共建重点 Venlo 型温室中进行,试验于 2012 年 9—11 月进行,10 月 1 日在江苏大学温室进行育苗,10 月 17 日选择长势、大小一致的生菜苗移栽到塑料盆中,每盆 1 株。

在成熟期进行农药喷洒试验,农药选用宜兴宜州化学制品有限公司生产的总有效成分含量为 20% 的氰戊菊酯乳油,分 4 种农药浓度,分别为 1.25 (高浓度)、0.83 (较高浓度)、0.6 (中浓度)、0.375 mL/L (低浓度),每种浓度农药喷洒 24 株生菜叶片,喷洒时保证均匀、湿透,共喷 96 株生菜,预留 12 株生菜未喷洒农药作为对照。

1.2 高光谱数据采集

采集不同处理生菜叶片,及时送至室内采集叶片光谱数据。光谱测量设备采用美国 ASD 公司生产的 FieldSpec® 3 型便携式光谱分析仪,该仪器光谱测量范围 350 ~ 2 500 nm,在 350 ~ 1 000 nm 光谱区采样间隔为 1.4 nm,分辨率为 3 nm;在 1 000 ~ 2 500 nm 光谱区采样间隔为 2 nm,分辨率为 10 nm。

将叶片放置于黑色绒布上,测量时将光谱探头置于工作台上方 5 cm 处,垂直于被测物,设定视场为 25°。测量叶片之前先测量标准反射板,以消除环境因素引起的系统误差,每片叶片测量 3 次,取平均值作为最终测量结果,不同浓度农药残留的生菜叶片原始光谱如图 1 所示。

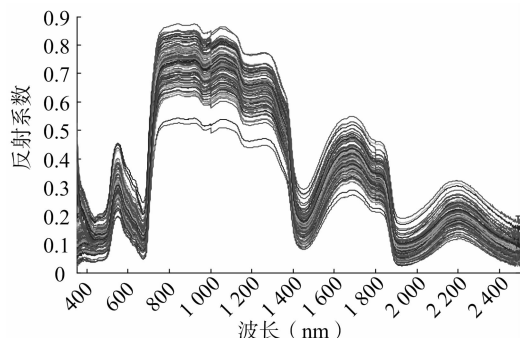


图1 不同浓度农药残留的生菜叶片的原始光谱

1.3 高光谱数据降噪

高光谱采集时,通常存在许多高频随机噪声,影响后续的光谱特征提取。本研究采用标准归一化 (SNV) 预处理方法,分别对每道光谱数据进行预处理,可以对基线漂移和光散射起到明显作用^[4]。经 SNV 处理后的光谱曲线如图 2 所示。

从图 2 中可以看出,在 500 nm 以下 2 400 nm 以上存在明显的噪声,在后期的数据处理中,选取 500 ~ 2 400 nm 范围内的光谱数据进行分析。

2 理论与算法

PCA、LPP、LDA、SLPP 4 种特征提取算法都可以被叫做子空间分析方法。通过寻找某个特定的投影矩阵,将原始的高维数据投影到低维的子空间中去,并满足算法所要求的某些

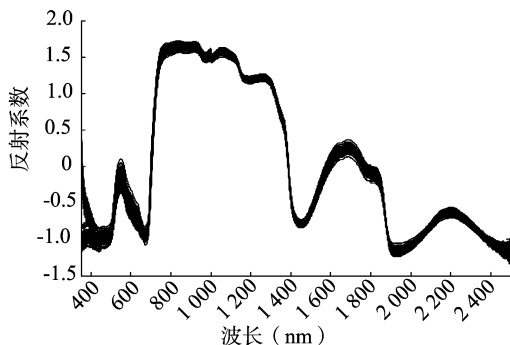


图2 SNV处理后的光谱曲线

特定条件,然后在低维子空间中对数据进行处理,不仅可以大幅度减少计算量,而且可以解决数据处理中常见的“维数灾难”问题。

2.1 降维算法

2.1.1 PCA 算法 PCA 是一种非监督的特征提取算法,主要思想是将数据空间经过正交变换映射到低维子空间的过程,这种变换在无损或者很少损失数据集信息的情况下降低了数据集的维数^[5]。

2.1.2 LDA 算法 LDA 是一种监督的特征提取算法,充分考虑了数据间的类别信息,通过寻找一个最优的投影矩阵,把高维数据投影到低维子空间中去,使得在子空间中,所有的样本数据达到类内离散度最小、类间离散度最大^[6],能使得投影后的数据更加有利于分类。

2.1.3 LPP 算法 LPP 是一种新的子空间分析算法,它是非线性方法 Laplacian Eigenmap 的线性近似,既能克服 PCA 等传统的线性方法难以保持非线性流形的缺点,又能解决传统的非线性方法难以获得新样本点低维投影的问题^[6]。

2.1.4 SLPP 算法 LPP 算法虽然可以保持数据之间的局部相似性关系,但它是一种非监督的特征提取算法,对于解决分类问题效果不一定好。可以通过引入监督机制将 LPP 改进为监督特征提取算法 SLPP^[7]。

2.2 SVM 分类算法

支持向量机是由 Vapnik 提出的一种新的机器学习方法,主要思想是通过一个非线性映射将低维线性不可分数据投影到高维空间使其线性可分,再在高维空间构建最优分类超平面对数据进行分类^[8]。

3 结果与分析

试验所用的生菜叶片样本共有 108 株,根据生菜叶片样本所喷洒的农药浓度的高低,将所有样本分为 5 类,分别为喷洒农药 1.250、0.830、0.600、0.375 mL/L 的生菜叶片样本各 24 个,未喷洒农药样本为 12 个生菜样本,在 Matlab 平台下进行试验仿真。

在喷农药的 4 类样本集中,分别从每类中挑出 18 个作为训练样本,剩下 6 个作为测试样本,在无农药的样本集合中,挑出 8 个作为训练样本,4 个作为测试样本。总训练样本 80 个,测试样本 28 个,然后进行建模试验。试验主要是为了 4 种特征提取算法的优缺点,并从中挑选出一种最适合于对生菜农药残留无损检测的算法。分别运行 PCA + SVM、LDA + SVM、LPP + SVM、SLPP + SVM 4 种算法,为了区分 4 种特征提

取算法的优劣,将 SVM 的核函数统一设为多项式内核函数,其他参数均为默认值。4 个算法分类结果分别如图 3、图 4、图 5、图 6 所示。

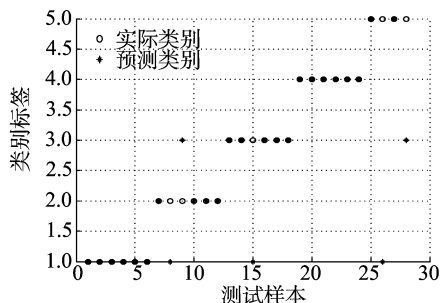


图3 PCA+SVM分类结果

从图 3 到图 6 可以发现,PCA + SVM 的分类正确率为 82.14%,LDA + SVM 的分类正确率为 89.29%,LPP + SVM 的分类正确率为 85.71%,SLPP + SVM 的分类正确率达到了 92.86%。这是由于有监督特征提取算法 LDA 和 SLPP 算法充分利用了数据类别信息,使降维后的数据对分类器更加敏

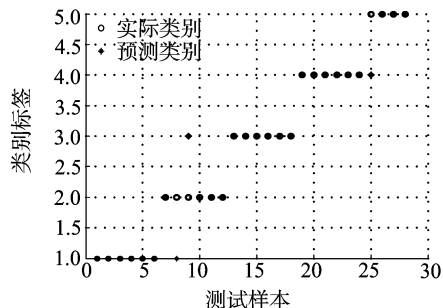


图4 LDA+SVM分类结果

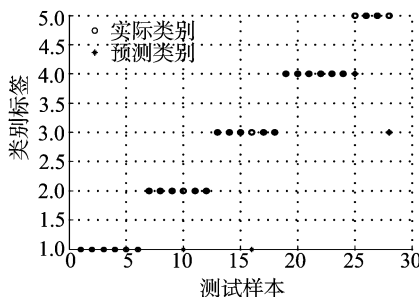


图5 LPP+SVM分类结果

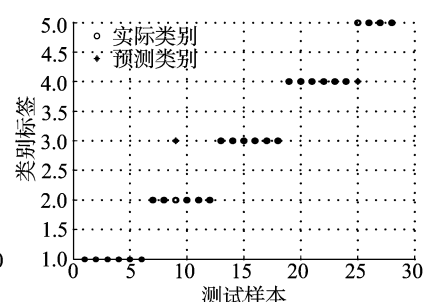


图6 SLPP+SVM分类结果

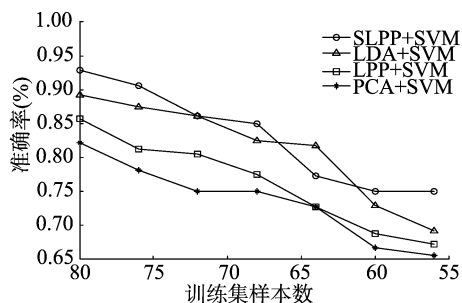


图7 4种算法在不同训练样本集下的分类准确率

4 结论

利用高光谱技术获得含有不同农药残留的生菜叶片样本的高光谱数据,采用标准归一化对高光谱数据进行预处理,通过观察预处理后的曲线,剔除明显噪声波段,然后分别运用 PCA、LDA、LPP、SLPP 4 种特征提取算法对数据进行处理,最后运用支持向量机对处理后的数据进行建模。结果表明,在解决分类问题上,与非监督 PCA 和 LPP 特征提取算法相比,监督特征提取算法 SLPP 和 LDA 具有明显优势。SLPP 能克服传统线性方法难以保持非线性流形的缺点,因此 SLPP + SVM 能达到较高的分类准确率。SLPP + SVM 可以用来作为生菜叶片农药残留无损检测的建模算法,为农业工程中关于农作物信息无损检测提供了一种新型有效解决途径。

感,分类效果更加明显。

为了更加有力证明本研究的观点,将训练集样本数从 80 个减少至 56 个,每次减少 4 个,测试集样本数对应地从 28 个增加到 52 个,4 种算法的分类正确率变化曲线如图 7 所示。

从图 7 可以看出,随着训练样本个数的不断减少,PCA + SVM、LDA + SVM、LPP + SVM 和 SLPP + SVM 的分类准确率均在不断下降,但是有监督特征提取算法 SLPP + SVM 和 LDA + SVM 2 条曲线始终位于非监督特征提取算法 PCA + SVM 和 LPP + SVM 曲线的上方,在解决分类问题时,有监督特征提取算法优于非监督特征提取算法。通过比较 SLPP + SVM 和 LDA + SVM 可以发现,当训练样本个数在 64 的时候,LDA + SVM 的分类正确率超过了 SLPP + SVM,其余时刻 SLPP + SVM 的分类准确率始终高于 LDA + SVM,本次试验中,SLPP 较 LDA 而言,更适用于对高光谱数据进行特征提取。因为 SLPP 能克服传统线性方法(LDA)难以保持非线性流形的缺点,利用 SLPP 进行特征提取,然后利用 SVM 进行分类建模能更好鉴别生菜叶片中农残浓度,为农作物农药残留无损检测提供了较为合适的方法。

参考文献:

- [1]陈蕊,张骏,李晓龙. 蔬菜表面农药残留可见-近红外光谱探测与分类识别研究[J]. 光谱学与光谱分析,2012,32(5): 1230-1233.
- [2]Sánchez M T, Flores - Rojas K, Guerrero J E, et al. Measurement of pesticide residues in peppers by near - infrared reflectance spectroscopy[J]. Pest Management Science,2010,66(6):580-586.
- [3]Soundar K R, Murugesan K. Preserving global and local information - a combined approach for recognising face images [J]. Computer Vision, IET,2010,4(3):173-182.
- [4]Xu L, Shi P T, Ye Z H, et al. Rapid geographical origin analysis of pure West Lake lotus root powder (WL - LRP) by near - infrared spectroscopy combined with multivariate modeling techniques [J]. Food Research International,2012,49(2):771-777.
- [5]李波,刘占宇,黄敬峰,等. 基于 PCA 和 PNN 的水稻病虫害高光谱识别[J]. 农业工程学报,2009,25(9):143-147.
- [6]He X F, Yan S C, Hu Y X, et al. Face recognition using laplacianfaces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2005,27(3):328-340.
- [7]申中华,潘永惠,王士同. 有监督的局部保留投影降维算法[J]. 模式识别与人工智能,2008,21(2):233-239.
- [8]薛宁静. 多类支持向量机分类器对比研究[J]. 计算机工程与设计,2011,32(5):1792-1795.