

单粒玉米种子成熟度快速判别方法

高彤¹ 吴静珠¹ 毛文华² 刘翠玲¹ 孙晓荣¹ 余乐¹

(1. 北京工商大学食品安全大数据技术北京市重点实验室, 北京 100048;
2. 中国农业机械化科学研究院, 北京 100083)

摘要: 为快速无损检测单粒种子成熟度,推进现代良种单粒播种技术,采用近红外光谱技术结合化学计量学方法建立了单粒玉米种子成熟度快速鉴别模型。采集 200 份单粒玉米种子样本近红外光谱,采用多种光谱预处理方法消除单粒种子采集过程中由于颗粒形态、杂散光等引入的干扰信息,分别采用分段谱区、连续投影算法筛选特征波长,优化建立种子成熟度判别模型。经 SG5_1 预处理后,在 5 500 ~ 4 000 cm⁻¹ 谱区建立的基于支持向量机的单粒玉米种子成熟度判别模型的分类准确率可达 92%。实验结果表明,近红外光谱技术可以快速无损判别单粒玉米种子成熟度。
关键词: 单粒玉米种子; 成熟度; 近红外光谱; 支持向量机
中图分类号: O657.33; S123 **文献标识码:** A **文章编号:** 1000-1298(2019)S0-0399-05

Method for Quickly Determining Maturity of Single Corn Seed

GAO Tong¹ WU Jingzhu¹ MAO Wenhua² LIU Cuiling¹ SUN Xiaorong¹ YU Le¹

(1. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China
2. Chinese Academy of Agricultural Mechanization Sciences, Beijing 100083, China)

Abstract: Seed maturity is an important factor affecting seed quality and crop yield. Therefore, rapid non-destructive testing of single seed maturity is an important basic technical guarantee for promoting modern single seed sowing technology for improved seeds. A rapid identification model of single corn seed maturity was established by combining near-infrared spectroscopy and chemometrics. The experiment collected the near infrared spectrum of 200 single corn seed samples. Interference information was reduced by particle morphology, stray light, etc. during single seed collection. The segmentation spectral region was used respectively, and the continuous projection algorithm was used to screen the feature wavelength optimization to establish a seed maturity determination model. The experiments showed that the recognition accuracy of the SVM based single corn seed maturity discrimination model established in the 5 500 ~ 4 000 cm⁻¹ spectral region could reach 92% after the Savitzky – Golay 5-point convolution first derivative pretreatment. The results showed that NIR had a bright application prospect in the rapid and non-destructive discrimination of single corn seed maturity.
Key words: single corn seed; maturity; near infrared spectroscopy; support vector machines

0 引言

玉米是我国重要的粮食作物、饲料作物和化工原料,加强玉米科技生产对于促进农业经济发展和保证国民经济稳定具有重要作用^[1-3]。单粒良种播种技术播种品质高、出苗整齐、节省种子、减少间苗

作业、降低成本及提高玉米整体效益等^[4-7],是玉米播种方式的栽培技术革新,是现代农业发展的必然趋势。因此,单粒良种播种技术对于玉米生产过程具有极为重要的意义。种子成熟度是影响种子品质的重要因素,成熟度低的种子其出芽率、芽势、拱土能力、成苗率和苗整齐度等综合指标下降,严

收稿日期: 2019-04-22 修回日期: 2019-05-25
基金项目: 国家重点研发计划项目(2018YFD0101004-03)
作者简介: 高彤(1995—),女,硕士生,主要从事分子光谱及光谱成像技术和农产品品质检测研究,E-mail: gaotonggz@163.com
通信作者: 吴静珠(1979—),女,副教授,博士,主要从事分子光谱及光谱成像技术和农产品品质检测研究,E-mail: pubwu@163.com

重影响玉米生产^[8-9]。因此单粒玉米种子成熟度的快速鉴别是玉米单粒良种播种技术推广的重要技术基础。

以近红外(Near infrared spectroscopy, NIRS)为代表的分子光谱及光谱成像技术以其非破坏性、快速测量的技术优势成为种子品质检测领域的研究热点^[10]。文献[11]研究不同成熟度哈密瓜样本在450~1 000 nm可见-近红外波段的光谱数据,建立哈密瓜成熟度判别模型,其成熟度判别正确率为95%。文献[12]研究400~1 000 nm波段范围的高光谱玉米籽粒成像技术,利用乳线区域面积占总体面积比例区分成熟度,找出区分度最高的特征波段图像,通过图像处理方法进行种子分类,结果表明:采用640 nm/525 nm的波段比图像,平均识别正确率为93.9%。文献[13]研究了380~1 023 nm波段范围内高光谱技术结合特征波长和光谱指数实现冬枣成熟度的可视化判别,结果表明:基于SPA和CARS选择的特征波长及引入的光谱指数分别建立的PLS-DA模型判别精度为97.27%、95.45%和98.18%。

上述研究表明,近红外光谱以及高光谱成像技术在建立种子及瓜果成熟度判别领域具有较好的应用效果。但是高光谱仪器检测成本相对较高,高光谱数据冗余量大,分析处理较为复杂。近红外技术相对于高光谱技术成本较为低廉,数据处理也较为简单,目前在烟草、奶制品、饲料^[14]等农副产品在线生产过程中的应用也日趋成熟和完善^[15-16]。本文提出采用近红外光谱技术结合化学计量学方法建立稳健的单粒玉米种子成熟度快速鉴别模型,探索近红外技术在单粒玉米种子成熟度快速鉴别领域的可行性,以期实现单粒种子成熟度近红外在线检测。

1 材料与方法

1.1 样本制备

实验玉米样本由种子市场购入,品种为中地77号和纯玉8号,两个品种各挑选100粒样本,共200粒。通常认为,成熟的种子是指种胚发育完全和种子后熟充足且具有良好发芽能力的种子。玉米完全成熟的直观标志有籽粒乳线消失,种子基部形成黑帽层(图1)等。实验中采用形成黑帽层作为划分种子成熟度的重要标志。每个品种根据基部是否出现黑层划分成熟样本50粒,未成熟样本50粒。根据K-S算法,样本按3:1比例划分训练集和测试集。实验样本为两个品种,训练集共150粒,测试集共50粒。



图1 成熟玉米种子与未成熟玉米种子对比
Fig.1 Comparison of mature corn seeds with immature corn seeds

1.2 近红外光谱采集

采用德国BRUKER公司的VERTEX70型傅里叶变换红外光谱仪(图2),扫描单粒玉米种子样本的近红外光谱。采用直径为2.5 cm的测量附件,籽粒胚向下放置,尽量保证每个籽粒扫描时,尖端方向一致(图3),以避免干扰。仪器参数设定如下:波数范围为12 500~4 000 cm⁻¹,分辨率为8 cm⁻¹,扫描64次,采样点数为2 074。扫描中地77号和纯玉8号样本,扫描得到近红外原始光谱如图4所示。



图2 VERTEX70型傅里叶变换红外光谱仪与附件
Fig.2 VERTEX70 Fourier transform infrared spectrometer and accessories



图3 玉米种子摆放位置

Fig.3 Corn seed placement

1.3 数据处理方法

1.3.1 连续投影算法

连续投影算法(Successive projections algorithm, SPA)是一种使矢量空间共线性最小化的前向变量

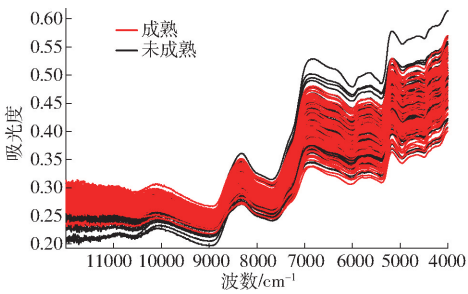


图4 单粒玉米种子近红外原始光谱

Fig. 4 Near-infrared raw spectrum of single-grain corn seeds

选择算法，它的优势在于提取全波段的几个特征波长，能够消除原始光谱矩阵中冗余的信息，可用于光谱特征波长的筛选^[17-18]。

1.3.2 支持向量机

支持向量机(Support vector machines, SVM)是一种二分类模型,其基本思想是:通过非线性变换将输入空间变换到一个高维的特征空间,并在新空间中寻找最优的线性分界面^[19-20]。目的是寻找一个超平面来对样本进行分割,分割的原则是间隔最大化,最终转换为一个凸二次规划问题来求解。

表 1 经过不同预处理后的近红外模型分类准确率

Tab. 1 Comparison of accuracy of near-infrared models after different pretreatments

光谱预处理方法	<i>c</i>	<i>g</i>	迭代次数	支持向量数	训练集分类准确率/%	测试集分类准确率/%
MSC	512	0.001 381 07	1 519	78	86.00	82
NOR	1 024	0.005 524 27	4 515	70	88.67	80
SG5_1	8	0.000 976 563	218	124	93.33	92
SNV	724.077	0.000 976 563	1 728	78	86.67	82

由表1得出,原始光谱数据经过SG5_1预处理后,有效消除了基线和其他背景的干扰,分辨重叠峰,提高了分辨率和灵敏度,此预处理方法效果最优。选取经SG5_1预处理后的数据建模,训练集分类准确率为93.33%,测试集分类准确率为92%。

2.2 基于分段谱区的单粒种子成熟度模型优化

在对近红外光谱进行解析时,由于近红外光谱主要是分子基频振动的倍频和合频吸收以及费米共振产生的吸收峰,通常根据分子化学键的基团不同,将其分成3个谱区:谱区I(12 500~8 500 cm⁻¹)、谱区II(8 500~5 500 cm⁻¹)和谱区III(5 500~4 000 cm⁻¹)。经过SG5_1预处理后的近红外光谱波段划分如图5所示,谱区II及谱区III单独建模;12 500~8 500 cm⁻¹的光谱噪声较大,故不单独建模;联合5 500~4 000 cm⁻¹和8 500~5 500 cm⁻¹两个噪声干扰小的波段,建立模型。

样本经过SG5_1预处理后,分段谱区建模结果

2 结果分析

2.1 不同预处理方法的模型对比

采用Matlab 2014a软件进行数据处理及建模。颗粒样品的漫反射光谱极易受样品状态和装样条件的影响。样品粒径是影响光谱测量的一个重要参数,随着样品粒径的增加,所测光谱的重现性变差,光谱的变动性随粒径的增加呈指数形式增加。每颗玉米籽粒形状、直径都不相同,所测光谱差异性变大,造成测量误差无法消除。因此需要采用光谱预处理方法进行光谱校正,消除干扰。实验选用多元散射校正(Multiplicative scatter correction, MSC)、标准正态变量变换(Standard normal variate transformation, SNV)、归一化(Normalization, NOR)、Savitzky-Golay卷积求导法(SG5_1)方法^[18]分别进行数据预处理。

鉴于单粒种子光谱采集过程中受非线性因素干扰较多,因此采用SVM建立判别模型。采用传统径向基函数RBF作为核函数,采用网格搜索法进行参数寻优,其中核参数*c*、*g*的取值范围均设置为[0.000 977, 1 024]。模型的分类结果如表1所示。

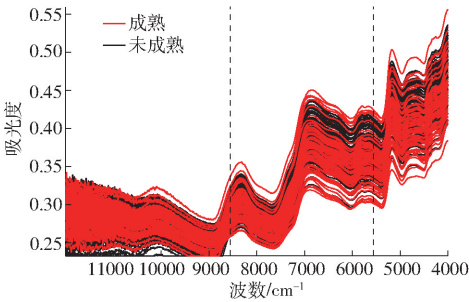


图5 单粒玉米种子近红外光谱波段划分

Fig. 5 Near-infrared spectral band division plot of single corn seed

如表2所示。

由于支持向量的个数影响训练与预测的速度,支持向量个数越多,训练和预测的速度越慢,5 500~4 000 cm⁻¹的模型参数*g*最大,支持向量数相对较少,训练速度更快,模型更稳定,无过拟合现象,建模效果较好,故本实验选取5 500~4 000 cm⁻¹的数据建模,训练集分类准确率96.67%,测试集分类准确

表 2 经 SG5_1 处理后的分段谱区模型分类准确率

Tab.2 Comparison of accuracy of segmented spectral region models after SG5_1

波数/cm ⁻¹	<i>c</i>	<i>g</i>	迭代次数	支持向量数	训练集分类准确率/%	测试集分类准确率/%
5 500 ~4 000	8	0. 062 5	229	104	96. 67	92
8 500 ~5 500	8	0. 002 762 14	200	121	88. 00	96
8 500 ~4 000	8	0. 000 976 563	120	113	94. 67	96
全光谱	8	0. 000 976 563	218	124	93. 33	92

率 92% ,惩罚系数为 8,*g* 为 0. 062 5。
5 500 ~ 4 000 cm⁻¹ 的主要吸收谱带有 C—H (CH₃,CH₂) 的组合频 $\nu + \delta$ 、游离 OH 和 C=O 的 2 倍频;6 710 ~ 6 500 cm⁻¹ 的吸收峰主要吸收谱带为游离 NH 的 1 倍频;7 090 ~ 6 900 cm⁻¹ 的吸收峰表征的是 C—H (CH₃,CH₂) 的组合频 $2\nu + 2\delta$;8 500 ~ 8 200 cm⁻¹ 的吸收峰主要表征的是 C—H (CH₃,CH₂) 的 2 倍频和组合频 $2\nu + 2\delta$ 。其波段主要吸收谱带是蛋白质和游离水的主要组成成分,故蛋白质和水分含量是区分玉米种子成熟度的重要因素。
由表 2 可以看出,8 500 ~ 5 500 cm⁻¹和 8 500 ~ 4 000 cm⁻¹ 的模型存在测试集准确率大于训练集准

准确率现象,这是由于训练集与测试集数据分布不一致,测试集样本比例较大导致。
为确保模型稳定性与可靠性,引入外部测试集,将 25 粒成熟玉米种子与 25 粒未成熟玉米种子组成外部测试集。测试集分类准确率为 91% ,与所建模型准确率接近,模型稳定。
2.3 基于 SPA 的单粒种子成熟度模型优化
为进一步简化模型,故选择 SPA 进行全光谱范围内特征波长提取。所选中的 4 个波长分别为 5 450. 123、5 608. 265、11 586. 82、8 767. 254 cm⁻¹。表 3 为样本原始光谱经过 SG5_1 预处理后,SPA 提取不同特征波长数所建模型分类准确率比较。

表 3 经 SG5_1 处理后的 SPA 模型分类准确率

Tab.3 Comparison of SPA model accuracy after SG5_1

选中波长数	<i>c</i>	<i>g</i>	迭代次数	支持向量数	训练集分类准确率/%	测试集分类准确率/%
4	45. 254 8	0. 250 000 0	134	61	84. 00	94
10	22. 627 4	0. 353 553 0	327	61	85. 33	88
149	22. 627 4	0. 007 812 5	541	124	74. 67	76

实验选取 4 个特定特征波长所建模型,模型迭代次数和支持向量数均减小,测试集分类准确率提高,达到 94%。表 3 中 3 个模型训练集分类准确率均低于测试集分类准确率,原因为全光谱经过 SPA 选取特征波长,其特征波长属于敏感特征点,虽然信息贡献率大,但是相较于全谱的 2 074 个波长点,仍不满足模型所需。提取特征波长建模是部分信息量的前提下简化模型。
5 450. 123 cm⁻¹ 和 5 608. 265 cm⁻¹ 处的近红外光谱主要吸收的是 CH₃ 和 CH₂ 的 1 倍频,主要是组成蛋白质的化学键;11 586. 82 cm⁻¹ 处表征的是游离 OH,主要是水分的组成;8 767. 254 cm⁻¹ 附近的吸收峰表征结合 OH,其中结合 OH 和蛋白质、糖类及磷脂等结合在一起。SPA 得到的 4 个波长处反映的相关基团是组成蛋白质和水分的化学键。由此可以看

出,区分玉米种子成熟度和蛋白质与水分含量有着密切关系。此模型结论与 5 500 ~ 4 000 cm⁻¹ 建模结论一致。

3 结束语

采用 SVM 方法结合近红外光谱建立单粒玉米种子成熟度鉴别模型,最终选取 SG5_1 预处理方法、5 500 ~ 4 000 cm⁻¹ 波段谱区建立单粒玉米种子成熟度的二分类定性模型,训练集分类准确率 96. 67% ,测试集分类准确率 92%。实验结果表明,玉米种子中的蛋白质及水分含量是影响单粒玉米种子成熟度的重要因素,将近红外光谱技术和模式识别方法相结合,用于单粒玉米种子成熟度判别具有可行性,可为单粒播种技术提供技术保障。

参 考 文 献

[1] 刘国梁,赵亚丽,王秀玲,等. 玉米种子成熟度对其活力及 F₁ 产量的影响[J]. 中国农业科学, 2016, 49(22):4342-4351. LIU Guoliang, ZHAO Yali, WANG Xiuling, et al. Effects of seed maturity on maize hybrid seed vigor and F₁ yield[J]. Scientia Agricultura Sinica, 2016, 49(22):4342-4351. (in Chinese)
[2] 陈龙江, MICHAEL R R. 种子质量对中国玉米产出的影响[J]. 华南农业大学学报(社会科学版), 2016, 15(3):19-27.

- CHEN Longjiang, MICHAEL R R. The effect of seeds quality improvement on maize yield[J]. Journal of South China Agricultural University (Social Science Edition), 2016, 15(3):19–27. (in Chinese)
- [3] 卢德成. 中国玉米生产成本优化研究[D]. 北京:中国农业科学院, 2018.
LU Decheng. Research on optimization of maize production cost in China[D]. Beijing:Chinese Academy of Agricultural Sciences, 2018. (in Chinese)
- [4] 戴立新. 玉米单粒播种高产技术推广优势和存在的问题[J]. 农业与技术, 2017, 37(18):55.
DAI Lixin. The advantages and problems of high yield technology of single seed sowing in maize[J]. Agriculture and Technology, 2017, 37(18):55. (in Chinese)
- [5] KAWASAKI M, KAWAMURA S, TSUKAHARA M, et al. Near-infrared spectroscopic sensing system for on-line milk quality assessment in a milking robot[J]. Computers and Electronics in Agriculture, 2008, 63(1):22–27.
- [6] CHEN L, YANG Z, HAN L. A review on the use of near-infrared spectroscopy for analyzing feed protein materials[J]. Applied Spectroscopy Reviews, 2013, 48(7):509–522.
- [7] 杨泉女, 周权驹, 蒙高山, 等. 近红外反射光谱技术测定甜玉米总多糖和淀粉含量的研究[J]. 吉林农业大学学报, 2019, 41(1):108–114.
YANG Quannü, ZHOU Qianju, MENG Gaoshan, et al. Determination of total polysaccharide and starch content in sweet corn by NIRS[J]. Journal of Jilin Agricultural University, 2019, 41(1):108–114. (in Chinese)
- [8] 卢德成. 中美玉米种植成本比较与分析[J]. 农业展望, 2018, 14(3):37–42.
LU Decheng. Comparison and analysis of maize planting cost between China and America[J]. Agricultural Outlook, 2018, 14(3):37–42. (in Chinese)
- [9] 朱占侠. 夏玉米单粒播种技术要点[J]. 种子科技, 2018, 36(7):44–45.
ZHU Zhanxia. Technology of single seed sowing of summer maize[J]. Seed Science & Technology, 2018, 36(7):44–45. (in Chinese)
- [10] 张鹏, 李江阔, 冯晓元, 等. 可见/近红外漫反射光谱预测磨盘柿成熟度[J]. 食品研究与开发, 2013, 34(11):91–94.
ZHANG Peng, LI Jiangkuo, FENG Xiaoyuan, et al. Prediction of maturity of mopan persimmon using near infrared diffuse reflectance spectroscopy[J]. Food Research and Development, 2013, 34(11):91–94. (in Chinese)
- [11] 孙静涛. 基于光谱和图像信息融合的哈密瓜成熟度无损检测研究[D]. 石河子:石河子大学, 2017.
SUN Jingtao. Study on nondestructive detection of hami melon maturity based on information fusion of spectrum and image[D]. Shihezi:Shihezi University, 2017. (in Chinese)
- [12] 杨小玲, 由昭红, 成芳. 高光谱成像技术检测玉米种子成熟度[J]. 光谱学与光谱分析, 2016, 36(12):4028–4033.
YANG Xiaoling, YOU Zhao hong, CHENG Fang. Study on identification of immature corn seed using hyperspectral imaging [J]. Spectroscopy and Spectral Analysis, 2016, 36(12):4028–4033. (in Chinese)
- [13] 曹晓峰, 任惠如, 李幸芝, 等. 高光谱技术结合特征波长/光谱指数对冬枣成熟度可视化判别[J]. 光谱学与光谱分析, 2018, 38(7):2175–2182.
CAO Xiaofeng, REN Huiru, LI Xingzhi, et al. Discrimination of winter jujube's maturity using hyperspectral technique combined with characteristic wavelength and spectral indices[J]. Spectroscopy and Spectral Analysis, 2018, 38(7):2175–2182. (in Chinese)
- [14] GIORGIO M, LORENZO S, ELISABETTA G, et al. Near-infrared calibration transfer for undried whole maize plant between laboratory and on-site spectrometers[J]. Italian Journal of Animal Science, 2018, 17(1):66–72.
- [15] 刘燕德, 徐海, 孙旭东, 等. 西红柿成熟度的近红外漫透射光谱无损检测[J]. 激光技术, 2019, 43(1):25–29.
LIU Yande, XU Hai, SUN Xudong, et al. Non-destructive measurement of tomato maturity by near-infrared diffuse transmission spectroscopy[J]. Laser Technology, 2019, 43(1):25–29. (in Chinese)
- [16] 蒋大鹏, 张冬妍, 李丹丹, 等. 基于近红外的松子蛋白质品质分类处理[J]. 计算技术与自动化, 2018, 37(3):180–184.
JIANG Dapeng, ZHANG Dongyan, LI Dandan, et al. Classification of pine nut protein quality based on near infrared[J]. Computing Technology and Automation, 2018, 37(3):180–184. (in Chinese)
- [17] 高云飞, 付霖宇, 瞿军, 等. 基于相似性度量的改进 KS 算法对近红外光谱分析模型的影响研究[J]. 电光与控制, 2019, 26(6):18–21.
GAO Yunfei, FU Linyu, QU Jun, et al. Research on influence of improved KS algorithm based on similarity measure on near-infrared spectroscopy analysis model[J]. Electronics Optics & Control, 2019, 26(6):18–21. (in Chinese)
- [18] 章海亮, 罗微, 刘雪梅, 等. 应用遗传算法结合连续投影算法近红外光谱检测土壤有机质研究[J]. 光谱学与光谱分析, 2017, 37(2):584–587.
ZHANG Hailiang, LUO Wei, LIU Xuemei, et al. Measurement of soil organic matter with near infrared spectroscopy combined with genetic algorithm and successive projection algorithm[J]. Spectroscopy and Spectral Analysis, 2017, 37(2):584–587. (in Chinese)
- [19] 孙丽萍, 张希萌, 何睿, 等. 基于 SVM 的近红外黑木耳多糖含量分类[J]. 电子科技, 2019(8):1–7.
SUN Liping, ZHANG Ximeng, HE Rui, et al. Near-infrared scanning polysaccharide content classification of auricularia Auricular based on SVM[J]. Electronic Science and Technology, 2019(8):1–7. (in Chinese)
- [20] TEYE E, AMUAH C L Y, MCGRATH T, et al. Innovative and rapid analysis for rice authenticity using handheld NIR spectrometry and chemometrics[J]. Spectrochimica Acta. Part A, Molecular and Biomolecular Spectroscopy, 2019, 217(3):147–154.