

基于CC-MPA特征优选算法的小麦条锈病遥感监测

竞霞 闫菊梅 邹琴 李冰玉 杜凯奇

(西安科技大学测绘科学与技术学院, 西安 710054)

摘要: 为了弥补一次性建模分析的缺陷,提高小麦条锈病遥感监测模型的运行效率和精度,根据模型集群分析(Model population analysis,MPA)算法的特点,综合利用光谱区间选择算法和光谱点选择算法的优势,提出了一种联合相关系数(Correlation coefficient,CC)与MPA的特征变量优选算法。在利用CC算法对全波段光谱进行特征变量选择的基础上,分别利用基于MPA思想开发的竞争性自适应重加权采样法(Competitive adaptive reweighted sampling,CARS)和变量组合集群分析法(Variable combination population analysis,VCPA)进一步优选对小麦条锈病敏感的特征变量,并利用偏最小二乘回归(Partial least squares regression,PLSR)算法构建了小麦条锈病遥感监测的CC-CARS和CC-VCPA模型。结果表明:联合CC-MPA算法优选的特征变量构建的CC-CARS和CC-VCPA模型精度均高于CC、CARS和VCPA算法。3组验证集样本中,CC-CARS模型预测病情指数(Disease index,DI)与实测DI间的 R^2_V 较CC模型和CARS模型至少分别提高了6.78%和6.66%,RMSEV至少分别降低了15.31%和10.98%,RPD至少分别提高了18.08%和12.34%。CC-VCPA模型预测DI与实测DI间的 R^2_V 较CC模型和VCPA模型至少分别提高了9.58%和0.73%,RMSEV至少分别降低了20.78%和3.86%,RPD至少分别提高了26.22%和4.02%。基于CC-MPA的光谱特征优选算法是一种有效的特征选择方法,尤其是利用CC-VCPA方法选择的特征变量数更少,模型预测效果更好,研究结果对光谱特征优选及提高作物病害遥感监测精度具有重要的参考价值。

关键词: 小麦条锈病; 遥感监测; 特征优选; 模型集群分析; 变量组合集群分析法

中图分类号: TP79; S435.121.4⁺2 文献标识码: A 文章编号: 1000-1298(2022)09-0217-09 OSID:



Remote Sensing Monitoring of Wheat Stripe Rust Based on CC-MPA Feature Optimization Algorithm

JING Xia YAN Jumei ZOU Qin LI Bingyu DU Kaiqi

(College of Geomatics, Xi'an University of Science and Technology, Xi'an 710054, China)

Abstract: In order to make up for the defects of the one-time modeling analysis and improve the operation efficiency and accuracy of wheat stripe rust remote sensing detection model, based on the characteristics of model population analysis (MPA) algorithm and the advantages of spectral interval selection algorithm and spectral point selection algorithm, a feature variable selection algorithm was proposed, combining correlation coefficient (CC) and MPA. Based on the selection of feature variables by CC algorithm for the full band spectrum, competitive adaptive reweighted sampling (CARS) and variable combination population analysis (VCPA) developed based on MPA were used to further optimize the feature variables sensitive to wheat stripe rust, and partial least squares regression (PLSR) algorithm was used to construct CC-CARS and CC-VCPA models for remote sensing monitoring of wheat stripe rust. The results showed that the accuracy of CC-CARS and CC-VCPA models constructed by combining the feature variables selected by CC-MPA algorithm was higher than that of CC, CARS and VCPA algorithm. In the three groups of validation set samples, CC-CARS model compared with CC model and CARS model, the R^2_V between predicted disease index (DI) and measured DI was increased by at least 6.78% and 6.66%, RMSEV was decreased by at least 15.31% and 10.98%, and RPD was increased by at least 18.08% and 12.34%, respectively. Compared the CC-VCPA model with CC

收稿日期: 2021-09-29 修回日期: 2021-10-26

基金项目: 国家自然科学基金项目(42171394、41601467、52079103)

作者简介: 竞霞(1978—),女,副教授,主要从事农业定量遥感研究,E-mail: jingxiaxust@163.com

model and VCPA model, the R_v^2 between predicted DI and measured DI was increased by 9.58% and 0.73%, RMSEV was decreased by 20.78% and 3.86%, and RPD was increased by 26.22% and 4.02%, respectively. The spectral feature optimization algorithm based on CC - MPA was an effective feature selection method. In particular, the number of feature variables selected by CC - VCPA method was less and the model prediction effect was better. The research results had important reference value for spectral feature optimization and improving the accuracy of remote sensing monitoring of crop diseases.

Key words: wheat stripe rust; remote sensing monitoring; feature optimization; model population analysis; variable combination population analysis

0 引言

小麦条锈病是一种发生范围广、危害程度大的病害,严重影响小麦的产量和质量^[1]。高光谱遥感数据维数高、信息量丰富,能够更好地反映植物胁迫信息,被广泛应用于小麦条锈病的遥感监测^[1-3],但也存在数据冗余、波段间相关性高的问题^[4]。从丰富的光谱数据中提取敏感光谱特征能够减少数据间的冗余,抑制噪声干扰,提高作物病害的监测精度,然而不同的特征选择算法对作物病害遥感监测模型构建的精度和效率具有重要影响^[5]。目前常用的遗传算法(Generalized algorithm,GA)^[6-7]、粒子群优化算法(Particle swarm optimization,PSO)^[8]和离散粒子群(Discrete binary particle swarm optimization,DBPSO)算法^[9]等智能优化特征选择算法虽然能够较好地保留变量之间的组合优势,但是这类算法结构较复杂,输入参数多,耗时长,影响了模型的实用性^[10];无信息变量消除法(Uninformative variables elimination,UVE)^[5-6]和连续投影算法(Successive projection algorithm,SPA)^[11-13]能够降低变量之间的共线性,然而该方法稳定性较差,提取的波长可能存在干扰变量,影响模型的预测精度^[13]。除了这些基于波长点的特征变量选择算法外,还有基于波长区间的CC法^[7,9]、间隔偏最小二乘(Interval partial least square,iPLS)^[14]和后向区间间隔偏最小二乘(Backward interval partial least square,BIPLS)^[15]等特征选择算法,这类算法选出的波长为连续分布的波段,具有较好的解释性^[16],但相邻波长变量之间往往具有较强的共线性特征,存在冗余变量,建模时难以达到最佳结果^[17]。上述光谱特征选择方法主要以一次建模为基础,数据处理易受样本个数的影响^[18]。针对这个问题,文献[19-21]提出了基于MPA思想的特征选择方法,该方法能够最大限度地利用已有样本集的信息,在众多变量中鉴定有信息变量、无信息变量和干扰变量,通过分析变量之间的相互作用选出最优变量组合。基于MPA思想的特征选择算法弥补了一次性建模分析的缺点,

能够增强模型预测结果的准确性和可解释性。

波长区间算法和波长点算法在特征选择中各有优点,波长区间选择算法是一种粗略选择特征变量的方法,能够在去除无信息变量的同时不遗漏关键变量,波长点选择算法能够精准选择重要变量,在保留较少变量的同时,取得较好的建模效果^[22]。基于此,本文提出联合CC与MPA的特征变量优选算法。首先使用CC方法进行波长区间的粗略选择,去除无信息变量确定最优特征波长谱区。然后利用基于MPA思想的CARS法和VCPA法精准选择对小麦条锈病严重度敏感的特征变量,并在此基础上采用PLSR算法建立小麦条锈病遥感监测模型。最后将该方法与仅利用CC、CARS、VCPA特征选择结果和建模精度进行对比分析,评价联合CC与MPA的特征变量优选算法在小麦条锈病遥感监测中的有效性,以期为高光谱数据选取特征变量进行病虫害遥感监测提供参考。

1 数据来源与研究方法

1.1 试验设计

试验区位于河北省廊坊市中国农业科学院试验站(39°30'40"N,116°36'20"E),小麦品种为对条锈病敏感的铭贤169号,平均种植密度为113棵/m²。2018年4月9日采用质量浓度为90 mg/L的孢子溶液对小麦进行条锈病接种。试验区域小麦分为健康组(编号为A、D)和染病组(编号为B、C),每个试验组面积为220 m²,分为8个样方,即健康组和染病组各16个样方。

1.2 数据获取

1.2.1 冠层光谱测量

本次试验于2018年5月18日测定小麦条锈病不同病情严重程度下的冠层光谱数据。冠层光谱测量所使用仪器为ASD Field Spec 4型地物光谱仪,其光谱分辨率为3 nm,采样间隔1.4 nm,采样波长范围350~2 500 nm,测量时间为11:00—12:30,测量高度距离地面1.3 m,探头视场角25°,共设有62个采样点,每个采样点观测10次并对观测结果取平均作为该采样点的光谱数据,每次测量前后均用标准

BaSO₄参考板对冠层辐亮度数据进行校正。

1.2.2 病情指数调查

冠层病情指数调查采用5点取样法,在每个样方内选取对称的5点,每点约1 m²,随机选取30株小麦,分别调查其发病情况。病情严重度参照国家标准^[23]进行量化。单叶严重度分为9个梯度,即0、1%、10%、20%、30%、45%、60%、80%和100%的叶片病斑覆盖,分别记录各病情严重度的小麦叶片数,计算不同梯度测试群体的病情指数

$$D_i = \frac{\sum_{i=1}^t x(i)f(i)}{t \sum_{i=1}^t f(i)} \times 100\% \quad (1)$$

式中 D_i ——病情指数, %

i ——梯度

$x(i)$ ——各梯度级值

t ——最高梯度,取9

$f(i)$ ——各梯度叶片数

1.3 基于MPA思想的特征选择算法

MPA的主要思路是通过随机采样对样本集中的特征变量信息进行采样并建立子模型,然后利用子模型构建新的样本空间,通过模型的回归系数等参数对新的样本空间进行进一步统计分析,以此获得有用信息^[19]。基于该思想,本文选择了CARS和VCPA 2种特征选择算法进行比较。

CARS是一种综合蒙特卡洛采样法(Monte Carlo sampling, MCS)和PLSR算法的特征波长变量选择方法^[24]。该方法通过指数衰减函数(Exponentially decreasing function, EDF)和自适应加权重采样技术(Adaptive reweighted sampling, ARS)选取回归系数较大的波长和建模时均方根误差(RMSE)最小的变量组合^[25]。VCPA是考虑变量之间相互作用的一种特征波长选择算法,该方法采用二进制矩阵采样法(Binary matrix sampling, BMS),每个变量被选中的机会是相等的^[26],通过EDF确定保留波长的数量,缩小变量空间,运用PLSR计算剩余变量子集之间所有组合的RMSE并进行排序优选^[27]。两种特征选择算法流程见图1。

1.4 模型构建与评价方法

1.4.1 光谱数据集的划分

为了评价所建模型的可靠性和稳定性,本文分别选用随机划分(Random sampling, RS)法、Kennard-Stone(KS)法和光谱-理化值共生距离(Sample set partitioning based on joint x-y distance, SPXY)法对62个样本集按照3:1的比例划分训练集和验证集。各划分方法的原理如下:

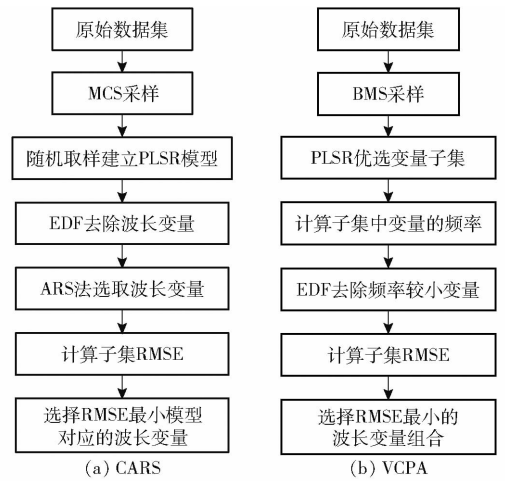


图1 基于MPA的特征选择算法流程图
Fig.1 Flowchart of feature selection algorithms based on MPA

(1) RS 法

RS法是随机选取一定数量的样本构成训练集和验证集,每个样本被选中的概率相同,无规律可循,能保证样本数据集的随机性^[28]。

(2) KS 法

KS法是按照样本光谱空间中的欧氏距离进行样本划分的方法。首先选择欧氏距离最大的2个样本构建数据集A,然后在剩余样本中选出与A样本中欧氏距离最大的样品加入到建模样品集,循环进行计算,直至挑选出足够样本数量的建模集,此时将剩余样本作为预测集样本^[29]。该方法优点是能保证训练库中样本按照空间距离分布均匀,缺点是需要进行数据转换和计算样本两两空间距离,计算量大。

(3) SPXY 法

SPXY法是一种同时考虑样本光谱信息以及生化参量欧氏距离的样本集选择方法,所选样本集更具有代表性^[30]。SPXY法与KS法的原理和步骤相似,主要区别是欧氏距离的计算公式不同。

$$d_{xy}(a,b) = \frac{d_x(a,b)}{\max_{a,b \in [1,n]} d_x(a,b)} + \frac{d_y(a,b)}{\max_{a,b \in [1,n]} d_y(a,b)} \quad (2)$$

其中 $d_x(a,b) = \sqrt{\sum_{j=1}^m (x_a(j) - x_b(j))^2}$ (3)

$$d_y(a,b) = |y_a - y_b| \quad (4)$$

式中 $d_{xy}(a,b)$ ——样本光谱信息与被测指标参量的样本a和b之间的距离

$d_x(a,b)$ ——样本a和样本b间x变量的欧氏距离

$d_y(a,b)$ ——样本a和样本b间y变量的欧

氏距离

- $x_a(j)$ ——样本 a 在第 j 个波长处的光谱信息值
- $x_b(j)$ ——样本 b 在第 j 个波长处的光谱信息值
- y_a ——样本 a 的被测指标参量
- y_b ——样本 b 的被测指标参量
- m ——样本光谱波长总数
- n ——样本总数

1.4.2 模型构建与评价

在对冠层光谱反射率矩阵与病情指数矩阵进行主成分分解的基础上利用 PLSR 算法构建小麦条锈病遥感监测模型。该算法是由多元线性回归分析、主成分分析和典型相关分析组合而成的算法,能够在自变量存在严重多重相关性的条件下进行有效回归建模,增加了预测准确性^[31]。

选择 DI 预测值和实测值之间的决定系数 R^2 、均方根误差 (RMSE) 和相对分析误差 (RPD) 对模型性能进行评价。对于同一样本集, R^2 越大, RMSE 越小, RPD 越大, 说明所建模型的预测能力和稳健性越高。其中, 当 RPD 大于等于 2 时, 表明模型可对

样本进行有效估测; 当 RPD 在 1.4 ~ 2 之间时, 表明模型可对样本进行粗略估测; 当 RPD 小于 1.4 时, 表明模型不能对样本进行估测^[32]。

2 结果与分析

2.1 基于 MPA 思想的特征变量选择

2.1.1 CARS 算法优选的特征变量

利用 CARS 算法优选特征变量过程如图 2 所示。由图 2a 看出, 随着采样次数的增加, 被选择的特征变量数逐渐减少, 结合图 2b、2c 可知, RMSE 在采样次数为 309 时最小, 表示此处冗余信息量最小, 之后随采样次数的增加开始缓慢增大, 这是因为此时光谱数据中的关键信息被剔除, 模型的性能变差, 因此选择 RMSE 最小值对应的图 3 中 22 个特征子集作为遥感监测小麦条锈病的最优特征变量。CARS 算法筛选出来的特征波长变量分别为 767、837、842、907、923、926、940、941、1 271、1 274、1 284、1 295、1 296、1 300、1 301、1 305、1 306、1 317、1 321、1 323、1 338、1 340 nm, 占总变量数的 2.191%。

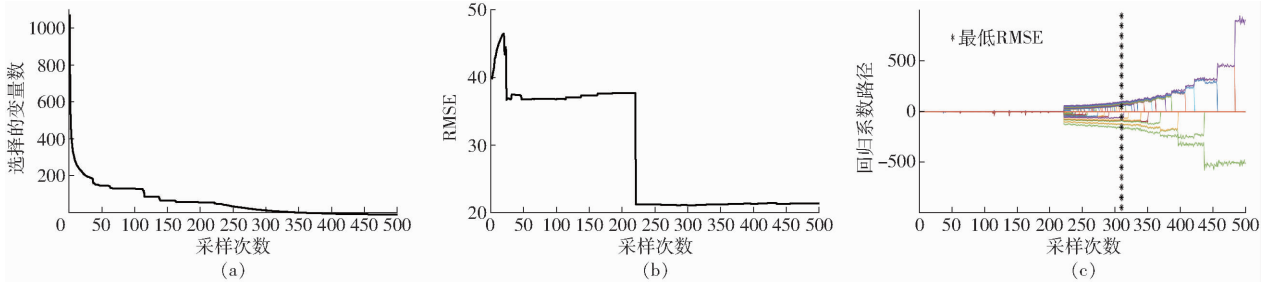


图 2 CARS 选择特征变量过程
Fig.2 Process of CARS selecting feature variables

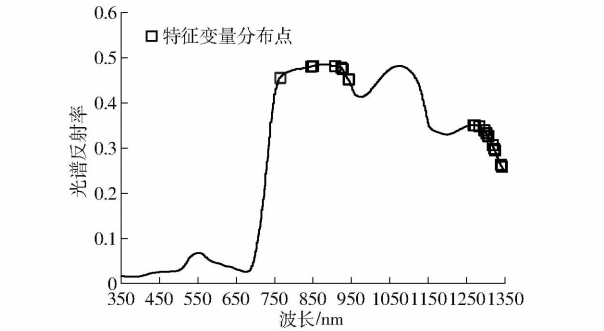


图 3 CARS 算法优选的特征变量
Fig.3 Feature variables selected by CARS algorithm

2.1.2 VCPA 算法优选的特征变量

采用 VCPA 算法优选特征变量时, 需要设置采样和迭代次数等参数, 各参数初始值设置如表 1 所示。首先利用 BMS 对光谱数据进行 1 000 次采样得到 1 000 组变量子集, 对这些变量子集建立 PLSR 模型, 从中选出 RMSE 最小的 100 组变量子集进行 EDF 运算, 最后建立模型计算保留下来的 14 个变量之间所有可能组合的 RMSE 值, 其中最小值对应的

表 1 VCPA 算法运行参数设置	
Tab.1 VCPA algorithm operation parameters setting	
参数	数值
BMS 采样次数 K	1 000
EDF 迭代次数 N	50
二进制矩阵 M 中每个变量被采样次数的占比 α	0.5
优秀子集占有比率 $\sigma/\%$	10
剩余变量数 L	14

变量即为最终选择的最优特征变量组合^[33]。采用 VCPA 优选的对小麦条锈病敏感的特征变量如图 4 所示, 分别为 415、508、513、528、533、647、648、803、1 022、1 288、1 295 nm, 共 11 个特征波长变量, 占总变量数的 1.099%。

2.1.3 CARS 和 VCPA 算法比较

为了评价 CARS 和 VCPA 2 种算法的稳定性, 以运行次数 50 为例, 对比分析了不同波长光谱被选为特征变量的频次 (图 5)。从图 5 可以看出, CARS 选择的变量较多, 涉及 297 个波长变量, 主要集中在

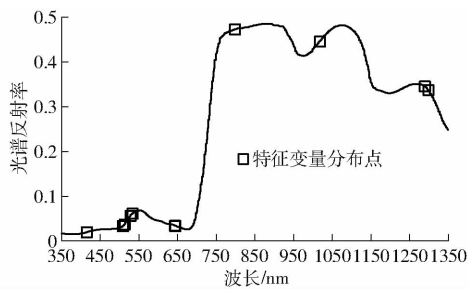


图 4 VCPA 算法优选的特征变量

Fig. 4 Feature variables selected by VCPA algorithm

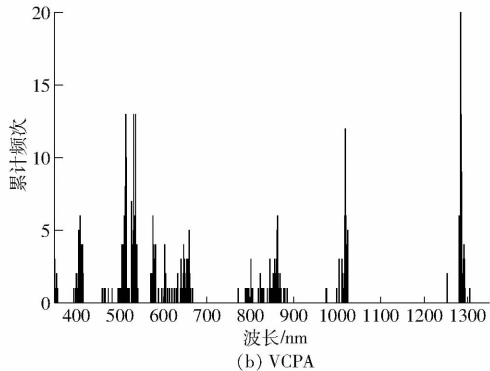
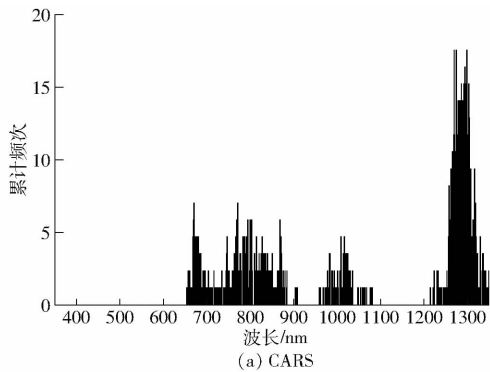


图 5 不同波长光谱被选为特征变量的频次

Fig. 5 Frequency of different wavelength spectra selected as feature variables

近红外波段 750 ~ 1 350 nm, 其中在波段 1 280 ~ 1 320 nm 范围内选择频次最高, 该波段是冠层叶片结构变化特征反映区^[34]。VCPA 选择的变量在全谱区域均有分布, 包含的信息更全面, 主要分布在 400 ~ 420 nm 叶绿素吸收带, 500 ~ 550 nm 反映类胡萝卜素的黄绿光反射区, 640 ~ 670 nm 叶绿素 b 吸收峰范围, 还有 800 ~ 870 nm、1 000 ~ 1 030 nm 和 1 270 ~ 1 330 nm 受叶片含水量影响的近红外波段, 该算法选择的变量比 CARS 少, 共涉及到 177 个波长变量, 稳定性更好。产生这种现象的原因是 2 种算法的采样策略不同, CARS 采用的 MCS 采样策略随机选择变量, 倾向于选择较多的变量, 每个变量被选择的机会是不确定的^[35]。而 VCPA 采用的 BMS 采样策略, 为每个变量提供相同的采样概率, 通过 EDF 压缩变量空间, 消除无关变量, 采用 MPA 思想保留了前 10% 最优的变量子集, 重

要变量被最终选择的几率相对较大, 所以最终保留的变量较少^[36]。

2.2 基于 CC 方法优选的特征变量

CC 是反映变量之间线性相关程度的统计指标, 数值越高表明该变量对于待测目标属性越重要。本文利用 Pearson 相关系数分析法计算小麦条锈病 DI 与冠层光谱反射率的相关性(图 6)。由图 6 可以看出, 可见光 577 ~ 703 nm 和近红外 735 ~ 962 nm、1 001 ~ 1 099 nm 波段内的光谱反射率与小麦条锈病严重度的相关性分别达到了极显著正相关和极显著负相关, 这是因为色素吸收决定着可见光波段的光谱反射率, 细胞结构决定近红外波段的光谱反射率。随着条锈病菌侵染程度的加重, 小麦叶片褪绿变黄, 叶绿素被大量破坏, 叶片细胞尺寸、形态、结构组织均发生变化, 水分蒸腾量显著增加^[37]。选择出与小麦条锈病 DI 达到极显著相关的波段区间作为小麦冠层条锈病病情监测模型的输入变量, 共 454 个波长特征变量, 占总变量数的 45.355%。

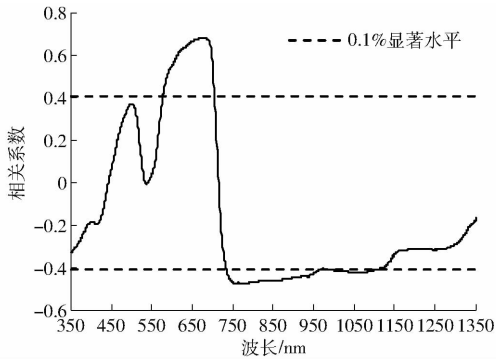


图 6 DI 与冠层光谱反射率的相关系数

Fig. 6 Correlation coefficient between DI and canopy spectral reflectance

2.3 模型构建与精度评价

在进行小麦条锈病严重度模型构建和精度评价时, 为了使评价结果更客观, 本文分别采用 RS 法、KS 法和 SPXY 法将 62 个样本集按 3 : 1 的比例进行 3 次分组, 每组中 45 个数据作为训练样本用于模型构建, 剩余的 17 个数据作为验证样本用于模型评价。

2.3.1 CARS 和 VCPA 模型

利用 CARS 和 VCPA 算法提取对小麦条锈病严重度敏感的特征变量, 基于 PLSR 算法构建了小麦条锈病遥感监测模型, 并将其与全波段光谱和 CC 方法提取的特征变量构建的模型进行比较, 以确定合适的特征变量选择方法, 表 2 显示了 3 组样本集中各模型的建模精度, 其中 R^2_T 和 R^2_V 分别表示训练集和验证集的决定系数, RMSET 和 RMSEV 分别表示训练集和验证集的均方根误差。

表 2 不同特征选择方法 PLSR 建模精度

Tab.2 PLSR modeling accuracy of different feature selection methods

特征变量 选择方法	变量数	RS 法					KS 法					SPXY 法				
		R^2_T	RMSET	R^2_V	RMSEV	RPD	R^2_T	RMSET	R^2_V	RMSEV	RPD	R^2_T	RMSET	R^2_V	RMSEV	RPD
全波段	1 001	0.842	0.105	0.714	0.199	1.817	0.918	0.088	0.697	0.163	1.610	0.917	0.074	0.698	0.210	1.804
CC	454	0.875	0.094	0.725	0.195	1.851	0.927	0.083	0.786	0.154	1.706	0.931	0.067	0.709	0.204	1.853
CARS	22	0.747	0.133	0.761	0.190	1.902	0.792	0.139	0.786	0.146	1.794	0.760	0.125	0.746	0.196	1.933
VCPA	11	0.811	0.115	0.870	0.141	2.570	0.899	0.097	0.800	0.153	1.718	0.845	0.101	0.823	0.174	2.177

从表 2 可以看出,利用 CARS 和 VCPA 2 种算法优选的特征变量数较全波段光谱有了极大减少,筛选的对小麦条锈病严重度敏感的特征变量分别为 22 和 11 个。在 3 组验证集样本中,与全波段的 PLSR 模型相比,基于 CARS 算法的 R^2_V 分别提高了 6.59%、12.87% 和 6.76%,RMSEV 分别降低了 4.49%、10.25% 和 6.66%,RPD 分别提高了 4.70%、11.42% 和 7.14%;基于 VCPA 算法的 R^2_V 分别提高了 21.94%、14.87% 和 17.91%,RMSEV 分别降低了 29.29%、6.27% 和 17.13%,RPD 分别提高了 41.42%、6.69% 和 20.67%。与 CC 法的 PLSR 模型相比,基于 CARS 算法的 R^2_V 分别提高了 4.90%、0.11% 和 5.18%,RMSEV 分别降低了 2.72%、4.87% 和 4.11%,RPD 分别提高了 2.79%、5.11% 和 4.29%;基于 VCPA 算法的 R^2_V 分别提高了 20.01%、1.89% 和 16.17%,RMSEV 分别降低了 27.97%、0.65% 和 14.86%,RPD 分别提高了 38.84%、0.66% 和 17.46%。由此可见 CARS 和 VCPA 2 种算法都能在减少变量的同时提升模型预测效果。

在 3 组样本集中,利用全波段建立的小麦条锈病遥感监测模型,其训练集样本的 R^2_T 均显著高于同组验证样本集的 R^2_V ,这是因为全波段光谱变量中存在干扰信息或无效变量,导致模型过于复杂,出现过拟合现象。基于 CC 法构建的模型精度较全波段光谱具有一定的提高,3 组样本集中,其 R^2_T 平均提高了 2.10%,RMSET 平均降低了 8.33%, R^2_V 平均提高了 5.28%,RMSEV 平均降低了 3.38%,RPD 平均提高了 3.53%,但仍然存在过拟合现象,说明采用 CC 法对模型的改善不明显。而基于 CARS 和 VCPA 挑选

特征波长变量后建立预测模型,其验证集 R^2_V 均高于全波段模型和 CC 模型,且与训练集的 R^2_T 较为接近,说明所建模型稳定性较好,减少了特征变量的数量,提高了模型的效率和预测精度。

对比 CARS 和 VCPA 2 种变量选择算法的表现,VCPA 选择的变量数更少,同时模型的预测结果更优。在 3 组样本集中,与 CARS 建模精度相比,VCPA 所建模型的 R^2_T 平均提高了 11.14%,RMSET 平均降低了 21.26%, R^2_V 平均提高了 8.88%,RMSEV 平均降低了 10.91%,RPD 平均提高了 14.48%。这是因为 CARS 在每次 MCS 采样后利用 EDF 函数强制删除了可能存在特征信息的次要变量与贡献较少的变量,导致模型拟合效果下降^[35],而 VCPA 的 BMS 采样策略为每个变量提供相同的选择机会,生成多个不同的子集并建立各种子模型,基于 MPA 思想最终保留了大部分关键变量。与 CARS 相比,VCPA 可以在保留重要变量的同时剔除更多的无用和冗余变量,所以 VCPA 模型对病情指数的预测效果更好,且倾向于选择更少的变量,泛化性能较好^[36]。

2.3.2 CC-CARS 和 CC-VCPA 模型

将基于波长区间粗选的 CC 方法和基于波长点精选的 CARS 和 VCPA 方法联用,能够充分利用它们的优势,精准选择重要变量,达到更好的建模效果。首先对全波段光谱进行 CC 优选,去除相关性较小的变量,将基于 CC 法选择的变量作为新变量利用 CARS 和 VCPA 进一步进行筛选,并利用 PLSR 算法分别建立了小麦条锈病遥感监测的 CC-CARS 和 CC-VCPA 模型,各模型预测 DI 和实测 DI 之间的 R^2 、RMSE 和 RPD 如表 3 所示。

表 3 联合 CC 的不同特征选择方法 PLSR 建模精度

Tab.3 PLSR modeling accuracy of different feature selection methods combined with CC

特征变量 选择方法	变量数	RS 法					KS 法					SPXY 法				
		R^2_T	RMSET	R^2_V	RMSEV	RPD	R^2_T	RMSET	R^2_V	RMSEV	RPD	R^2_T	RMSET	R^2_V	RMSEV	RPD
CC-CARS	17	0.791	0.121	0.837	0.152	2.372	0.830	0.126	0.839	0.130	2.015	0.799	0.115	0.837	0.160	2.372
CC-VCPA	10	0.832	0.108	0.877	0.135	2.673	0.845	0.120	0.861	0.103	2.550	0.858	0.096	0.832	0.162	2.339

由表 3 可知,CC-CARS 和 CC-VCPA 分别选择了 17 个和 10 个波长变量,进一步减少了特征变

量数。相比于 CC、CARS 和 VCPA 模型,CC-CARS 和 CC-VCPA 模型效果均得到提升,在 3 组验证样

本集中,CC-CARS方法建立的小麦条锈病监测模型比CC方法的 R^2_V 分别提高了15.44%、6.78%和18.14%,RMSEV分别降低了21.98%、15.31%和21.87%,RPD分别提高了28.17%、18.08%和28.00%;比CARS算法的 R^2_V 分别提高了10.05%、6.66%和12.32%,RMSEV分别降低了19.80%、10.98%和18.52%,RPD分别提高了24.69%、12.34%和22.73%。CC-VCPA方法建立的小麦条锈病监测模型比CC方法的 R^2_V 分别提高了20.89%、9.58%和17.45%,RMSEV分别降低了30.76%、33.08%和20.78%,RPD分别提高了44.42%、49.43%和26.22%;比VCPA算法的 R^2_V 分别提高了0.73%、7.56%和1.10%,RMSEV分别降低了3.86%、32.64%和6.94%,RPD分别提高了4.02%、48.46%和7.46%,说明将CARS、VCPA算法与CC法联用能够在减少模型复杂度的同时提高模型的稳定性和预测精度。其中基于CC-VCPA模型预测效果最佳,选择的变量数最少,3个样本集中训练集 R^2_T 分别为0.832、0.845和0.858,验证集 R^2_V 分别为0.877、0.861和0.832,RPD分别为2.673、2.550和2.339,均满足了建模精度,联合方法取得较好效果,它充分利用了波长区间选择的强解释性,减少无效波长点对波长精选算法的干扰,帮助波长精选在较小的区间中进行最优选择,不但能提升模型的解释能力和运算效率,还可以降低因参与变量过多而容易引发的过拟合现象的发生概率。

为了更直观地显示CC-MPA方法选择的特征变量在小麦条锈病遥感监测中的有效性,RS、KS和SPXY3组样本集中RMSEV和选择特征变量数(Number of variables, nVAR)从全波段到最优特征变量选择的变化趋势如图7所示。由图7可以看出,本文所提出的CC-MPA特征选择方法不断地将全波段变量空间从CC粗选收缩到最优选择CC-CARS与CC-VCPA,整体上RMSEV和nVAR都呈现下降的趋势,在由CC粗选到CC-CARS与CC-VCPA精选的过程中, RMSEV下降的速度变快, nVAR降低到最少,最后获得一个变量更少且精度更高的特征子集。同时CC-CARS和CC-VCPA分别与单次CARS和VCPA建模结果相比, nVAR和RMSEV均进一步减少,说明联合CC与MPA的特征优选方法能够有效收缩特征变量空间并提高预测结果精度。

3 讨论

与CC特征选择方法相比,基于MPA思想的CARS和VCPA算法能够更加有效地提取对小麦条

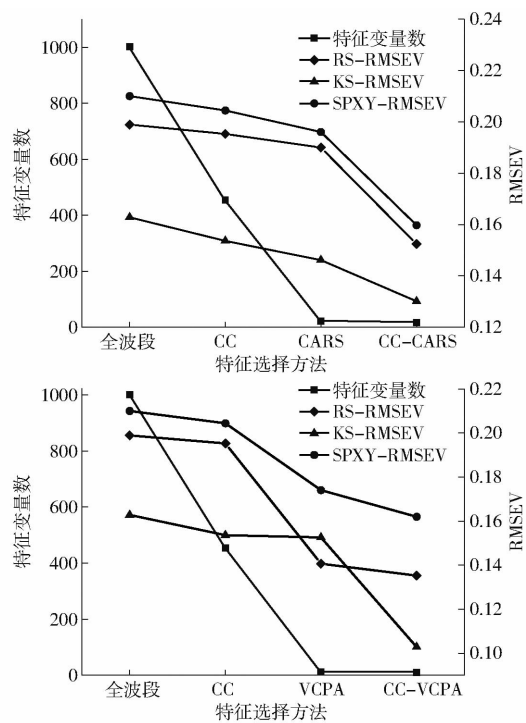


图7 不同方法选择的特征变量数(nVAR)和RMSEV的变化曲线

Fig.7 Variation trend of nVAR and RMSEV selected by different methods

锈病敏感的特征变量,提高小麦条锈病遥感监测精度和建模效率。这是因为基于MPA思想的特征选择方法能最大限度地利用已有样本集信息,更好地分析光谱变量间的组合效应,进而得到最优特征变量^[38]。利用CC法提取的特征波段建立模型的预测精度较低,主要原因是CC方法虽然在一定程度上可以剔除包含冗余信息的变量,但仍然存在保留变量多和筛选结果中存在较低信噪比变量等不足,导致模型容易出现过拟合现象^[39]。

利用CC-MPA算法优选的特征变量构建的小麦条锈病遥感监测模型精度优于单一的CARS和VCPA算法。这是因为联合CC与MPA的特征选择方法能够综合波长区间算法和波长点算法的优点,在利用CC算法对全波段进行粗选去除无信息变量的基础上,进一步利用MPA方法精选重要变量,保留较少变量的同时建立更稳健的模型,取得更好的建模效果。此外,利用CC-VCPA算法选择的特征变量数最少,构建的模型精度最高,这是因为CARS和VCPA2种算法的采样策略不同。CARS算法基于MCS策略随机选择变量,每个变量被选择的概率不同,不能充分评价每个变量的重要性,导致该算法选择的特征波长较多且相对比较集中,缺失了部分有用信息,多样性不足^[35],进而影响模型的预测准确性。而VCPA算法采用BMS采样策略,每个变量都有相同的被选择概率^[36],提取的变量在全谱区域

均有分布,信息量丰富,提高了特征波长的多样性,因此模型的精度较 CARS 算法有了明显提高。

本研究只选择基于 MPA 思想的 2 种特征选择算法进行分析,由于不同类型特征选择算法对变量筛选的原理不同,如果使用更多的特征选择算法进行试验,CC-VCPA 方法是否仍为遥感探测小麦条锈病的最优特征选择算法尚需进一步探讨。此外,为了充分发挥特征选择算法之间的互补性,在后续研究中可以将智能优化算法、SPA 算法等特征选择方法与本文方法联合进行特征变量的三重优选,以进一步探讨更优的光谱特征优选方法。

在分析 CC-MPA 方法提取小麦条锈病遥感监测的敏感因子时,本研究仅利用了原始反射率光谱数据,若增加能够敏感反映作物光合生理特征的叶绿素荧光数据及反映不同生化组分含量、冠层几何结构的植被指数特征,是否可以提高模型监测精度有待考证。

本研究的样本数据主要为小区控制试验条件下小麦灌浆期测定的近地高光谱数据,小麦品种、生育期以及其它病虫害等因素对冠层光谱的影响较小,因此该研究结果能否外推到不同光谱分辨率、不同空间分辨率的卫星影像上,还需利用更多的数据进一步分析。

4 结论

(1)为从全波段光谱数据中优选出对小麦条锈病敏感的特征变量,提高小麦条锈病遥感监测精度,

提出了联合 CC 与 MPA 思想的特征变量选择方法,对比分析 CC、CARS、VCPA 以及 CC-CARS、CC-VCPA 几种特征变量选择方法对小麦条锈病遥感监测精度的影响。

(2)CARS 和 VCPA 算法均能在减少特征变量的同时提升模型预测效果。CARS 和 VCPA 算法分别选择了 22 个和 11 个对小麦条锈病敏感的光谱特征,极大地减少了特征变量数,提高了模型运行效率。在 3 组样本集中,基于 CARS 和 VCPA 算法优选特征变量建立的小麦条锈病遥感监测模型的精度较全波段和 CC 方法均有不同程度的提高。与 CARS 相比,VCPA 算法选择的变量数更少,模型预测精度更高,表明 VCPA 算法可以有效选择对小麦条锈病敏感的特征变量,简化模型并提高运算效率,增强模型可靠性。

(3)联合 CC 与 MPA 的特征变量选择方法能够提高小麦条锈病遥感监测精度。将 CC 法与基于 MPA 思想开发的 CARS 和 VCPA 算法联合优选对小麦条锈病敏感的特征变量,并在此基础上构建的模型精度较单一的 CC、CARS 和 VCPA 模型均有不同程度的提高,其中以 CC-VCPA 模型预测结果最好,3 组样本集中验证模型预测 DI 和实测 DI 间的 RPD 均大于 2,表明 CC-MPA 是一种有效的特征变量优选算法。研究结果为高光谱特征变量优选及提高作物病害监测精度提供了研究方法和思路,具有重要的参考价值。

参 考 文 献

- [1] 王静,景元书,黄文江,等. 冬小麦条锈病严重度不同估算方法对比研究[J]. 光谱学与光谱分析,2015,35(6):1649-1653.
WANG Jing, JING Yuanshu, HUANG Wenjiang, et al. Comparative study on different estimation methods of severity of stripe rust of winter wheat[J]. Spectroscopy and Spectrum Analysis, 2015, 35(6):1649-1653. (in Chinese)
- [2] 黄文江,刘林毅,董莹莹,等. 基于遥感技术的作物病虫害监测研究进展[J]. 农业工程技术,2018,38(9):39-45.
HUANG Wenjiang, LIU Linyi, DONG Yingying, et al. Research progress of crop pest monitoring based on remote sensing technology[J]. Agricultural Engineering Technology, 2018, 38(9):39-45. (in Chinese)
- [3] 竞霞,吕小艳,张超,等. 基于 SIF-PLS 模型的冬小麦条锈病早期光谱探测[J]. 农业机械学报,2020,51(6):191-197.
JING Xia, LÜ Xiaoyan, ZHANG Chao, et al. Early detection of winter wheat stripe rust based on SIF-PLS model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(6):191-197. (in Chinese)
- [4] 竞霞,邹琴,白宗璠,等. 基于反射光谱和叶绿素荧光数据的作物病害遥感监测研究进展[J]. 作物学报,2021,47(11):2067-2079.
JING Xia, ZOU Qin, BAI Zongfan, et al. Research progress of crop diseases monitoring based on reflectance and chlorophyll fluorescence data[J]. Acta Agronomica Sinica, 2021, 47(11):2067-2079. (in Chinese)
- [5] 张东辉,赵英俊,秦凯,等. 光谱变换方法对黑土养分含量高光谱遥感反演精度的影响[J]. 农业工程学报,2018,34(20):141-147.
ZHANG Donghui, ZHAO Yingjun, QIN Kai, et al. Influence of spectral transformation methods on nutrient content inversion accuracy by hyperspectral remote sensing in black soil[J]. Transactions of the CSAE, 2018, 34(20):141-147. (in Chinese)
- [6] 刘燕德,肖怀春,孙旭东,等. 柑桔叶片黄龙病光谱特征选择及检测模型[J]. 农业工程学报,2018,34(3):180-187.
LIU Yande, XIAO Huaichun, SUN Xudong, et al. Spectral feature selection and discriminant model building for citrus leaf Huanglongbing[J]. Transactions of the CSAE, 2018, 34(3):180-187. (in Chinese)
- [7] 竞霞,张腾,白宗璠,等. 融合 GA 与 SVR 算法的小麦条锈病特征优选与模型构建[J]. 农业机械学报,2020,51(11):253-263.
JING Xia, ZHANG Teng, BAI Zongfan, et al. Feature selection and model construction of wheat stripe rust based on GA and SVR algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(11):253-263. (in Chinese)

- [8] 胡根生,吴问天,黄文江,等. 粒子群优化的最小二乘支持向量机在小麦白粉病监测中的应用[J]. 遥感技术与应用, 2017,32(2):299-304.
HU Gensheng, WU Wentian, HUANG Wenjiang, et al. Application of PSO-LSSVM in wheat powdery mildew monitoring[J]. Remote Sensing Technology and Application, 2017,32(2):299-304. (in Chinese)
- [9] 白宗璠,竞霞,张腾,等. MDBPSO 算法优化的全波段光谱数据协同冠层 SIF 监测小麦条锈病[J]. 作物学报, 2020,46(8):1248-1257.
BAI Zongfan, JING Xia, ZHANG Teng, et al. Canopy SIF synergize with total spectral reflectance optimized by the MDBPSO algorithm to monitor wheat stripe rust[J]. Acta Agronomica Sinica, 2020,46(8):1248-1257. (in Chinese)
- [10] GALVAO R K H, ARAUJO M C U, MARTINS M D, et al. An application of subagging for the improvement of prediction accuracy of multivariate calibration models[J]. Chemometrics and Intelligent Laboratory Systems, 2006,81(1):60-67.
- [11] CHENG S X, XIE C Q, WANG Q N, et al. Different wavelengths selection methods for identification of early blight on tomato leaves by using hyperspectral imaging technique[J]. Spectroscopy and Spectral Analysis, 2014,34(5):1362-1366.
- [12] 刘爽,谭鑫,刘成玉,等. 高光谱数据处理算法的小麦赤霉病籽粒识别[J]. 光谱学与光谱分析, 2019,39(11):3540-3546.
LIU Shuang, TAN Xin, LIU Chengyu, et al. Recognition of fusarium head blight wheat grain based on hyperspectral data processing algorithm[J]. Spectroscopy and Spectral Analysis, 2019,39(11):3540-3546. (in Chinese)
- [13] 王帅. 小麦赤霉病高光谱特征提取方法研究[D]. 合肥:安徽农业大学, 2019.
WANG Shuai. Research on extraction methods for hyperspectral characteristics of wheat scab[D]. Hefei: Anhui Agricultural University, 2019. (in Chinese)
- [14] LI X L, XU K W, ZHANG Y Y, et al. Optical determination of lead chrome green in green tea by fourier transform infrared (FT-IR) transmission spectroscopy[J]. Plos One, 2017, 12(1): e0169430.
- [15] 程术希. 基于光谱和成像技术的作物病害不同侵染期快速检测方法研究[D]. 杭州:浙江大学, 2014.
CHENG Shuxi. Fast detection methods for crop disease infection period using spectral and imaging technology[D]. Hangzhou: Zhejiang University, 2014. (in Chinese)
- [16] YU H D, YUN Y H, ZHANG W, et al. Three-step hybrid strategy towards efficiently selecting variables in multivariate calibration of near-infrared spectra[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2020, 224: 117376.
- [17] ZOU X B, ZHAO J W, POVEY M J W, et al. Variables selection methods in near-infrared spectroscopy[J]. Analytica Chimica Acta, 2010, 667(1-2): 14-32.
- [18] LI H D, LIANG Y Z, XU Q S, et al. Model population analysis for variable selection[J]. Journal of Chemometrics, 2010, 24(7-8): 418-423.
- [19] LING H D, ZENG M M, TAN B B, et al. Recipe for revealing informative metabolites based on model population analysis[J]. Metabolomics, 2010, 6(3): 353-361.
- [20] 李洪东. 广义灰色分析体系建模的基本问题及其模型集群分析研究[D]. 长沙:中南大学, 2012.
LI Hongdong. Fundamental problems of generalized grey analytical systems and their research of model population analysis[D]. Changsha: Central South University, 2012. (in Chinese)
- [21] LI H D, LIANG Y Z, CAO D S, et al. Model-population analysis and its applications in chemical and biological modeling[J]. Trends in Analytical Chemistry, 2012, 38(1): 154-162.
- [22] YUN Y H, LI H D, DENG B C, et al. An overview of variable selection methods in multivariate analysis of near-infrared spectra[J]. Trends in Analytical Chemistry, 2019, 113: 102-115.
- [23] 中国国家标准化管理委员会. GB/T 15795—2011 小麦条锈病测报技术规范[S]. 北京:中国标准出版社, 2011.
- [24] 石岩,孙冬梅,熊婧,等. 近红外光谱结合竞争性自适应重加权采样算法用于人工牛黄的质量分析研究[J]. 中国药杂志, 2018, 53(14): 1216-1221.
SHI Yan, SUN Dongmei, XIONG Jing, et al. Analysis of artificial cow-bezoar by near-infrared spectroscopy coupled with competitive adaptive reweighted sampling method[J]. Chinese Pharmaceutical Journal, 2018, 53(14): 1216-1221. (in Chinese)
- [25] WANG Y, JIANG F, GUPTA B B, et al. Variable selection and optimization in rapid detection of soybean straw biomass on CARS[J]. IEEE Access, 2018, 6: 5290-5299.
- [26] YUN Y H, WANG W T, DENG B C, et al. Using variable combination population analysis for variable selection in multivariate calibration[J]. Analytical Chimica Acta, 2015, 862: 14-23.
- [27] 宋雨宸,宦克为,韩雪艳,等. 基于蒙特卡洛变量组合集群分析法的小麦蛋白质近红外光谱变量选择[J]. 长春理工大学学报(自然科学版), 2017, 40(5): 29-35.
SONG Yuchen, HUAN Kewei, HAN Xueyan, et al. Variable selection of wheat protein near infrared spectral based on Monte Carlo variable cluster analysis[J]. Journal of Changchun University of Science and Technology (Natural Science Edition), 2017, 40(5): 29-35. (in Chinese)
- [28] 秦冲,陈雯雯,何雄奎,等. 近红外光谱分析中建模校正集的选择[J]. 光谱学与光谱分析, 2009, 29(10): 2661-2664.
QIN Chong, CHEN Wenwen, HE Xiongkui, et al. Study on a method of selecting calibration samples in NIR spectral analysis[J]. Spectroscopy and Spectral Analysis, 2009, 29(10): 2661-2664. (in Chinese)
- [29] 黄慧,张德钧,詹舒越,等. 干贝水分检测的建模及分级方法[J]. 光谱学与光谱分析, 2019, 39(1): 185-192.
HUANG Hui, ZHANG Dejun, ZHAN Shuyue, et al. Research on sample division and modeling method of spectrum detection of moisture content in dehydrated scallops[J]. Spectroscopy and Spectral Analysis, 2019, 39(1): 185-192. (in Chinese)

- traceability system based on blockchain[J]. Transactions of the CSAE, 2019, 35(22): 323–330. (in Chinese)
- [25] 祝烈煌, 董慧, 沈蒙. 区块链交易数据隐私保护机制[J]. 大数据, 2018, 4(1): 46–56.
ZHU Liehuang, DONG Hui, SHEN Meng. Blockchain transaction data privacy protection mechanism[J]. Big Data, 2018, 4(1): 46–56. (in Chinese)
- [26] 祝烈煌, 高峰, 沈蒙, 等. 区块链隐私保护研究综述[J]. 计算机研究与发展, 2017, 54(10): 2170–2186.
ZHU Liehuang, GAO Feng, SHEN Meng, et al. A review of blockchain privacy protection research[J]. Computer Research and Development, 2017, 54(10): 2170–2186. (in Chinese)
- [27] 韩璇, 袁勇, 王飞跃. 区块链安全问题: 研究现状与展望[J]. 自动化学报, 2019, 45(1): 206–225.
HAN Xuan, YUAN Yong, WANG Feiyue. Blockchain security issues: research status and prospects[J]. Acta Automatica Sinica, 2019, 45(1): 206–225. (in Chinese)
- [28] 王群, 李馥娟, 王振力, 等. 区块链原理及关键技术[J]. 计算机科学与探索, 2020, 14(10): 1621–1643.
WANG Qun, LI Fujuan, WANG Zhenli, et al. Blockchain principles and key technologies[J]. Computer Science and Exploration, 2020, 14(10): 1621–1643. (in Chinese)
- [29] 贺海武, 延安, 陈泽华. 基于区块链的智能合约技术与应用综述[J]. 计算机研究与发展, 2018, 55(11): 2452–2466.
HE Haiwu, YAN An, CHEN Zehua. Overview of smart contract technology and application based on blockchain[J]. Computer Research and Development, 2018, 55(11): 2452–2466. (in Chinese)
- [30] 谢瑞彬, 李会, 张娟, 等. 鲐毒鱼毒素检测及物种鉴定技术研究进展[J]. 食品科学, 2022, 43(5): 306–315.
XIE Ruibin, LI Hui, ZHANG Juan, et al. Recent advances in techniques for tetrodotoxin detection and species identification of tetrodotoxic fish: a review[J]. Food Science, 2022, 43(5): 306–315. (in Chinese)
- [31] 朱燕超. 面向区块链系统的查询处理研究[D]. 上海: 华东师范大学, 2020.
ZHU Yanchao. Research on query processing for blockchain systems[D]. Shanghai: East China Normal University, 2020. (in Chinese)
-

(上接第 225 页)

- [30] GALVAO R K H, ARAUJO M C U, JOSE G E, et al. A method for calibration and validation subset partitioning[J]. Talanta, 2005, 67(4): 736–740.
- [31] 罗批, 郭继昌, 李镭, 等. 基于偏最小二乘回归建模的探讨[J]. 天津大学学报(自然科学与工程技术版), 2002, 35(6): 783–786.
LUO Pi, GUO Jichang, LI Qiang, et al. Modeling construction based on partial least squares regression[J]. Journal of Tianjin University (Science and Technology), 2002, 35(6): 783–786. (in Chinese)
- [32] 秦占飞, 常庆瑞, 谢宝妮, 等. 基于无人机高光谱影像的引黄灌区水稻叶片全氮含量估测[J]. 农业工程学报, 2016, 32(23): 77–85.
QIN Zhanfei, CHANG Qingrui, XIE Baoni, et al. Rice leaf nitrogen content estimation based on hyperspectral imagery of UAV in Yellow River diversion irrigation district[J]. Transactions of the CSAE, 2016, 32(23): 77–85. (in Chinese)
- [33] 孙宗保, 王天真, 刘小裕, 等. 高光谱结合波长选择算法串联策略检测调理牛排新鲜度[J]. 光谱学与光谱分析, 2020, 40(10): 3224–3229.
SUN Zongbao, WANG Tianzhen, LIU Xiaoyu, et al. Detection of prepared freshness using hyperspectral technology combined with wavelengths selection methods combination strategy[J]. Spectroscopy and Spectral Analysis, 2020, 40(10): 3224–3229. (in Chinese)
- [34] 黄文江. 作物病害遥感监测机理与应用[M]. 北京: 中国农业科学技术出版社, 2009: 102–150.
- [35] JI H, WANG W Z, CHONG D F, et al. CARS algorithm-based detection of wheat moisture content before harvest[J]. Symmetry, 2020, 12(1): 115.
- [36] JIANG H, XU W D, CHEN Q S. Comparison of algorithms for wavelength variables selection from near-infrared (NIR) spectra for quantitative monitoring of yeast (*Saccharomyces cerevisiae*) cultivations[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 214: 366–371.
- [37] 刘良云. 植被定量遥感原理与应用[M]. 北京: 科学出版社, 2014: 136–178.
- [38] 宋相中, 唐果, 张录达, 等. 近红外光谱分析中的变量选择算法研究进展[J]. 光谱学与光谱分析, 2017, 37(4): 1048–1052.
SONG Xiangzhong, TANG Guo, ZHANG Luda, et al. Research advance of variable selection algorithm in near infrared spectroscopy analysis[J]. Spectroscopy and Spectral Analysis, 2017, 37(4): 1048–1052. (in Chinese)
- [39] 刘潭, 许童羽, 于丰华, 等. 改进特征波段选取和混合集成建模的东北粳稻叶绿素含量估算[J]. 光谱学与光谱分析, 2021, 41(8): 2556–2564.
LIU Tan, XU Tongyu, YU Fenghua, et al. Chlorophyll content estimation of northeast japonica rice based on improved feature band selection and hybrid integrated modeling[J]. Spectroscopy and Spectral Analysis, 2021, 41(8): 2556–2564. (in Chinese)