doi:10.6041/j.issn.1000-1298.2021.02.018

基于改进分离阈值特征优选的秋季作物遥感分类

王庚泽[1]　靳海亮[1]　顾晓鹤[2,3]　杨贵军[3,4]　冯海宽[3,4]　孙　乾[2,3]

(1. 河南理工大学测绘与国土信息工程学院, 焦作 454000;
2. 北京农业信息技术研究中心农业农村部农业遥感机理与定量遥感重点实验室, 北京 100097;
3. 国家农业信息化工程技术研究中心, 北京 100097; 4. 北京市农业物联网工程技术研究中心, 北京 100097)

摘要: 为了提高秋季作物分类精度,以多时相的 Sentinel-2 为数据源,以生育进程相近的秋季作物为分类对象,提出一种基于 Relief F 算法和信息熵改进分离阈值算法(Modified ISEaTH-based entropy, EMISE)的多评价准则融合特征优选算法——改进分离阈值组合式特征优选算法(Modified EMISE-based Relief F, ReEMISE),并分析了不同特征对秋季作物分类的重要性。首先,利用 Relief F 算法对特征进行初选,结合 EMISE 算法对 2 种评价准则进行融合,再优化初选特征集,进而利用随机森林(Random forest ,RF)方法提取农作物种植面积,并与单评价准则的 Relief F 算法和 EMISE 算法的随机森林分类精度进行比较。同时,利用多时相光谱特征、传统指数特征、红边指数特征、纹理特征、不同时相波段差值特征、不同时相波段比值特征及优选特征,通过 7 组不同的特征组合提取秋季作物种植面积,分析不同特征组合对秋季作物分类精度的影响。结果表明:ReEMISE 特征优选的随机森林法在特征变量为 9 个时精度最高,总体精度和 Kappa 系数分别为 95.391 8% 和 0.939 7;综合多特征是提高农作物分类精度的关键,在多时相光谱特征基础上分别加入传统指数特征和红边特征,总体精度分别提高 1.502 1、1.571 5 个百分点,Kappa 系数分别提高 0.019 8、0.020 7。因此综合多特征的 ReEMISE 特征优选的随机森林法可以有效提高秋作物分类精度和效率。

关键词: 秋季作物; 遥感分类; 特征优选; 改进分离阈值组合式特征优选算法; 随机森林

中图分类号: S127　　**文献标识码**: A　　**文章编号**: 1000-1298(2021)02-0199-12　　**OSID**:

Remote Sensing Classification of Autumn Crops Based on Hybrid Feature Selection Model Combining with Relief F and Improved Separability and Thresholds

WANG Gengze[1]　JIN Hailiang[1]　GU Xiaohe[2,3]　YANG Guijun[3,4]　FENG Haikuan[3,4]　SUN Qian[2,3]

(1. *School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo* 454000, *China*
2. *Key Laboratory of Quantitative Remote Sensing in Agriculture, Ministry of Agriculture and Rural Affairs, Beijing Research Center for Information Technology in Agriculture, Beijing* 100097, *China*
3. *National Engineering Research Center for Information Technology in Agriculture, Beijing* 100097, *China*
4. *Beijing Engineering Research Center for Agriculture Internet of Things, Beijing* 100097, *China*)

Abstract: The multi-temporal Sentinel-2 images were used to classify the autumn crops in Gaocheng, Shijiazhuang to provide an important basis for the local agricultural planting structure adjustment. The influence of comprehensive multi-features and feature optimization on the extraction accuracy of autumn crop planting area were analyzed. In order to reduce the influence of high-dimensional features on the performance of the classifier, a filter hybrid feature selection model (ReEMISE) based on improved separability and thresholds combined with the Relief F algorithm was proposed. Firstly, the Relief F dimensionality reduction algorithm was used to select the features. Secondly, the improved separability and thresholds (EMISE) combined with image entropy was used to further optimize the preliminary feature set, and then the EMISE feature importance value was given a Relief F feature weight. Finally,

收稿日期: 2020-04-23　修回日期: 2020-06-04

基金项目: 国家自然科学基金项目(41571323)和国家重点研发计划项目(2016YFD0300609)

作者简介: 王庚泽(1994—),男,研究实习员,主要从事农业定量遥感研究,E-mail: wanggengze18@126.com

通信作者: 顾晓鹤(1979—),男,研究员,博士,主要从事农业遥感相关研究,E-mail: guxh@nercita.org.cn

the random forest model was used to extract the crop planting area from the optimized feature subset, and compared with the random forest classification accuracy of the Relief F dimensionality reduction algorithm and the EMISE dimensionality reduction algorithm. The purpose was to ensure the accuracy of classification, minimize the feature dimensions and improve the classification efficiency. Six different types of feature variables were generated based on the Sentinel-2 data with multi-phase and rich spectral information, including multi-temporal spectral features, traditional index features, red-edge index features, texture features, difference features of different time-phase bands, and ratio features of different time-phase bands. On the basis of multi-temporal spectral features, adding different features, totally six groups of different feature combination experiments were constructed to extract autumn crop planting area and verify the classification accuracy of different feature combination. At the same time, the influence of different features on the extraction accuracy of crop planting area was analyzed from two aspects: the importance of features and the classification accuracy of different feature combinations. The results showed that the best correlation coefficient threshold of the ReEMISE's feature-optimized random forest model (RF_ReEMISE) was 0.96. The accuracy was the highest when the number of feature variables was 9 and the overall accuracy and Kappa coefficient were 95.391 8% and 0.939 7, respectively. The advantages of the RF_ReEMISE mainly included two aspects, i.e., the least number of feature variables at the highest accuracy, and the classification accuracy index of ReEMISE algorithm was the highest when the number of different feature variables of the three algorithms attained the best accuracy. Comprehensive multi-feature was the key to improve the accuracy of crop classification. Based on the multi-temporal spectral features, the traditional index features and the red-edge index features were added respectively, and the overall accuracy was improved by 1.502 1 percentage points and 1.571 5 percentage points, and the Kappa coefficient was increased by 0.019 8 and 0.020 7. The accuracy was slightly reduced when adding the texture features, the difference features of different time-phases bands, and the ratio features of different time-phase bands. The RF_ReEMISE can reduce the feature dimensions, improve the classification accuracy, and achieve the balance between efficiency and accuracy. The spectral differences of crops at different phenological stages can be better reflected by introducing the ratio features of different time-phase bands than the difference features of different time-phase bands. It was found that correlation between the red-edge vegetation index and chlorophyll was stronger, and the importance of features was higher. The RF_ReEMISE with multi-features can effectively improve the accuracy and efficiency of autumn crop classification.

Key words: autumn crops; remote sensing classification; feature optimization; ReEMISE; random forest

0 引言

快速准确获取农作物种植面积是农情监测、农业种植结构调整和农业经济政策制定的重要基础数据[1-3]。传统的逐级上报和抽样调查方法,具有一定的主观因素,费时费力、时效性差。利用遥感技术可以及时、准确、客观、快速获取大范围农作物的分布,是目前农业遥感的研究热点之一[4-5]。然而,农作物之间光谱相似,具有“同物异谱、异物同谱”现象。尤其秋季作物的种植结构复杂、种植周期相同、生育期较短,同时受秋季云雨天气较多和卫星回访周期影响,难以获得高质量和最佳物候期的遥感数据[6]。目前,国内外学者主要从时相、特征以及分类方法等角度进行了大量研究[3]。

随着模式识别技术在遥感图像分类方面的应用,国内外学者一直在探索高精度、高效率的遥感图像分类方法。早期利用多时相光谱信息和农作物物候信息提取农作物时,大多数学者采用决策树方法[7],但传统的决策树法需人工设定阈值。随着机器学习技术的发展,有学者引入支持向量机[8-9]和人工神经网络[10-11]等方法来提取农作物。由于随机森林方法不容易出现过拟合,对噪声有一定的容忍度,且分类精度高,近年来在农作物提取上得到广泛的应用[12-13]。

利用以上方法提取秋季作物时,综合多时相遥感的多特征信息提取与优化是目前研究的热点之一。结合光谱特征、光谱衍生的指数特征、时相特征、纹理特征和 DEM 等多特征变量,使其优势互补,可以有效提高分类精度[14]。文献[15]通过 RapidEye 卫星影像,比较了红边光谱特征和纹理特征在提取韩国早熟型、中熟型和中晚熟型 3 种水稻种植面积中的重要性,发现在单时相中红边特征能较好提高不同品种水稻的分类精度,纹理特征只带来微小的精度提高。文献[7]将 DEM 数据作为特征参与决策树分类,有效提高了山区农作物分布的位置精度和面积精度,解决了山区分类精度比平原

差的问题。文献[16]为提取黑龙江省北安市大豆、水稻、春小麦和玉米的空间分布信息，在 Sentinel－2A 多时相光谱的基础上，引入不同农作物、不同物候期的 NDVI 时序变化特征，对比发现，总体精度和 Kappa 系数分别提高 7.7% 和 0.055。

特征过多容易降低分类器性能，造成精度和效率降低[17]。多特征优选可最大程度降低特征维数，使分类器具有较好的泛化能力。优选的特征集应具有特征和目标相关性强、特征之间冗余小和类间可分性强的特点[18]。目前和随机森林分类器结合最多的是 Relief F 算法和基于随机森林的平均不纯度减少算法。文献[19]利用 Landsat－8 OLI 数据提取黄河口滨海湿地信息，发现利用 Relief F 特征优选的随机森林法提取湿地信息具有高精度、高效率的优势。文献[20]提出一种多目标遗传随机森林组合式特征选择算法(MOGARF)，结合随机森林分类法提取湿地信息，并分别与 Relief F 算法、Boruta 算法和完整特征集进行对比，结果表明，总精度提高了 0.35% ~ 1.94%，代外误差降低 0.91% ~ 1.48%，特征维度降低至 10%。文献[21]综合多特征采用平均不纯度减少算法优化特征子集，特征重要性排序前 9 个特征参与随机森林分类时精度最高，为热带、亚热带复杂土地覆盖分类提供了借鉴经验。以上 2 种特征优选算法在一定程度上提高了分类精度，但算法上仍有不足，平均不纯度减少算法特征关联导致分类贡献高特征可能重要性得分低，Relief F 算法没有考虑特征之间的相关性。

多特征组合和特征优选是遥感图像分类研究的热点之一。目前，基于特征优选的随机森林分类方法应用于秋季作物的分类研究较少。Relief F 算法是经典的 Filter 式特征选择算法，但其随机性对数量不均衡的高维小样本具有一定的局限性。EMISE 算法充分考虑类内和类别之间的分离度，针对每种地物提取分离度高的特征。本文针对单一降维算法特征评价不全面的问题，通过取长补短将 Relief F 算法和 EMISE 算法融合，以增加类间距离、减小类内距离为基本评价准则，提出一种结合 Relief F 算法的改进分离阈值组合式特征优选算法(ReEMISE)，并结合随机森林法实现对石家庄市藁城区秋季作物种植面积的提取。同时分析不同特征组合对秋季作物分类精度的影响和不同特征对秋季作物分类的重要性，通过 ReEMISE 降维算法与 Relief F 算法和 EMISE 算法的对比，以验证 ReEMISE 特征优选的随机森林分类方法(RF_ReEMISE)在秋季作物分类中的优势。

1　研究区与数据获取

1.1　研究区概况

藁城区位于河北省西南部(37.51° ~ 38.18°N，114.39° ~ 114.59°E，见图 1)，石家庄市东侧，总面积 836 km^2，耕地约 540 km^2。由于土壤多由洪积冲积物、河流冲积物组成，肥沃可耕、灌溉便利，是华北平原的主要产粮区，素有“河北粮仓”、“燕赵天府”之称。该区秋季作物主要以小麦、玉米一年两熟的轮作模式为主，其中东南部与赵县接壤，以贾市庄镇为主，主要种植梨树；与晋州区接壤的常安镇是大棚蔬菜的主要种植基地；滹沱河北部以南董镇的北洼和北高庄村为中心，是花卉苗木种植基地。近年来，藁城区为适应市场需求积极引导农民调整种植结构，大力帮扶农民种植大豆和富硒谷子。该区秋季作物以玉米和大豆为主，小宗农作物包括谷子、花生和高粱。秋季作物在 6 月中上旬播种，8 月进入旺盛期，9 月下旬至 10 月上旬为收获期。

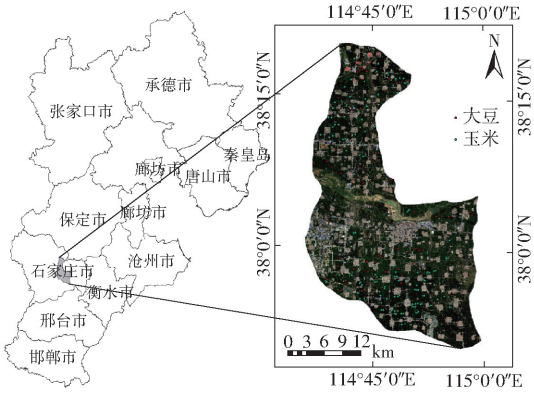

图 1　研究区位置

Fig. 1　Location of study area

1.2　数据获取与预处理

Sentinel－2 数据下载自欧洲航天局(European Space Agency，ESA)的数据共享网站(https：//scihub. copernicus. eu/)。Sentinel－2 分别由欧洲航天局于 2015 年 6 月和 2017 年 3 月发射的双星组成，其双星重访周期 5 d，最高空间分辨率 10 m。其波段从可见光到短波红外包括 13 个光谱波段，是唯一包含 3 个红边波段(中心波长分别为 560、665、705 nm)的光学遥感卫星。根据研究区作物物候特征，选取无云质量高的 Sentinel－2 上层大气反射的 L1C 级影像 3 景，成像时间分别为 2019 年 7 月 2 日、2019 年 8 月 16 日和 2019 年 8 月 31 日。采用精度最高[22]的 SNAP 软件中的大气校正模型 Sen2cor 对下载数据进行大气校正，对影像的 10 个波段(不包括 b1、b9 和 b10)双线性重采样为空间分辨率 10 m 的ENVI 格式的影像数据。2019 年 9 月底开展

详细的藁城实地农业种植调查,采用高精度的手持 GPS 获取经纬度信息并拍照记录。根据实地调查情况,影像地物分为 5 大类,包括玉米、大豆、经济林(果林、景观林)、小宗作物(谷子、花生、高粱)和非植被,主要秋季作物玉米、大豆的分布见图 1。其中玉米、大豆、经济林、小宗作物和非植被样本数分别为 204、198、121、114、189 个,按照 2∶1 随机分为训练样本和验证样本。

2 研究方法

2.1 特征说明

目前,综合多时相遥感影像的多特征变量提取与优化是农作物遥感识别的研究热点之一。Sentinel-2 影像丰富的光谱信息具有植被红边波段的独特优势,研究发现通过引入红边波段数据及其相关指数,能整体提高农作物面积识别精度[23-25],纹理特征丰富的地物信息,也可以提高中高分辨率遥感影像的分类精度[1-2]。王娜等[26]研究发现不同时相遥感影像的波段差值在优选特征子集中占大约 35%。因此,本文选择原始光谱、传统光谱指数、红边指数、不同时相波段差值、不同时相波段比值和纹理 6 类特征,共 183 个特征。

原始光谱包括 3 景影像的 10 个波段(b2、b3、b4、b5、b6、b7、b8、b8A、b11、b12)的光谱反射率,共 30 个特征。

不同时相波段差值和不同时相波段比值是任意组合两组不同时相遥感影像对应波段之间的差值或比值[26],共 3 组影像,30 个不同时相波段差值特征和 30 个不同时相波段比值特征。在 Python 3.7 中利用 GDAL(Geospatial data abstraction library)批处理得到。

传统植被指数[4]与红边植被指数[27-28]区别在于是否有红边波段参与计算。3 景影像共 27 个传统植被指数和 42 个红边植被指数。

根据前人研究,通过对原始影像主成分分析的第 1 主成分,在 ENVI 5.3 中利用灰度共生矩阵提取纹理特征[4],主要包括对比度(contrast)、熵(entropy)、自相关(correlation)、均匀性(mean)、同质性(homogeneity)、方差(variance)、角二阶矩(Second moment)、非相似性(dissimilarity)。3 景影像共 24 个纹理特征。

2.2 特征优选方法

2.2.1 Relief F 算法原理

KIRA 等[29] 1992 年提出了计算特征权重的 Relief 算法。其基本思想通过最近邻学习,随机选取一个样本,较好的特征应该是同类样本之间的距离较小,而不同类样本之间的距离较大。通过赋予每个特征的权重判断特征的重要性。但 Relief 仅适用 2 个类别的样本。KONONENKO[30] 提出了多类特征选择的 Relief F 算法。具体实现为:在样本中随机抽取一个实例样本 R,在实例样本 R 的同类样本中搜索 K 个最近邻样本,分别在实例样本 R 所有的不同类样本中搜索 K 个最近邻样本,组成同类近邻样本集合 H 和不同类近邻样本集合 T。再通过 R 与 H 和 T 中样本的距离计算特征权重 W。假如 m 个特征,随机抽样 n 次的特征权重 W_m 的计算式为

$$W_{m(n)} = W_{m(n-1)} - \frac{\sum_{j=1}^{k} \frac{|R[m] - H_j[m]|}{\max(m) - \min(m)}}{nK} + \frac{\sum_{c \notin \text{class}(R)} \left(\frac{P(c)}{1 - P(\text{class}(R))} \sum_{j=1}^{k} \frac{|R[m] - T(c)_j[m]|}{\max(m) - \min(m)} \right)}{nK} \tag{1}$$

式中 c——异类样本的类别

$R[m]$——特征 m 中的随机样本 R 的值

$H[m]$——特征 m 中 R 的同类样本最近邻值

$T(c)[m]$——特征 m 中 R 的 c 类样本最近邻值

$\max(m)$——特征 m 中最大值

$\min(m)$——特征 m 中最小值

$\text{class}(R)$——样本 R 的类别

$P(c)$ 表示第 c 类目标的概率,其可由该类样本数与样本总数之比确定。

2.2.2 基于图像熵的分离阈值算法(EMISE)

NUSSBAUM 等[31] 2006 年提出了分离阈值(Separability and thresholds, SEaTH)算法。该算法通过 J-M(Jeffries-Matusita)距离判断类别在特征上的分离度,对于 2 个类别其距离 J 为

$$J = 2(1 - e^{-B}) \tag{2}$$

$$\text{其中} B = \frac{1}{8}(a_1 - a_2)^2 \frac{2}{\sigma_1^2 + \sigma_2^2} + \frac{1}{2}\ln\left(\frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1^2\sigma_2^2}\right) \tag{3}$$

式中 a_i、σ_i——某特征上类别的均值和标准差,$i = 1, 2$

B——巴氏距离

余晓敏等[32]考虑到特征之间的冗余,提出了去相关性和加权类内距离的改进的分离阈值算法(Improved SEaTH, ISEaTH)。通过 Pearson 相关系数,设置阈值去掉相关性强的特征,其加权类内距离 D 计算公式为

$$D = (k_1 d_1 + k_2 d_2)/(k_1 + k_2) \tag{4}$$

$$\text{其中} \quad d_i = \frac{2}{k_i(k_i - 1)} \sum_{l=1}^{k_i} \sum_{m=l+1}^{k_i} (f_j^l - f_j^m)^2 \quad (i = 1, 2) \tag{5}$$

式中　d_1、d_2——某特征上某类别内所有样本的距离累加和

k_1、k_2——某类样本数量

f_j^l、f_j^m——某类内样本和的特征值

ISEaTH 根据类间距离、类内距离和 Pearson 相关系数提取影像最优特征子集，但没有考虑图像信息量。杨惠雯等[33]引入图像信息熵，提出了一种结合图像信息熵的阈值选择方法。信息熵公式为

$$H(X) = -\sum_x p(x)\lg(p(x)) \tag{6}$$

式中　$p(x)$——相应的概率分布函数

根据类间距离大和类内距离小原则，结合图像信息熵构建类间可分性的评价指标计算式为

$$S = \frac{J}{D}H \tag{7}$$

可知 S 越大，类别的可分性越好。

2.2.3　ReEMISE 算法

Relief F 算法具有效率高，不受数据类型限制的优势[34]。但只考虑特征与类别之间的相关性，忽略了特征之间的冗余性，特征评价不全面。EMISE 算法不仅考虑图像的信息量，还通过 Pearson 相关系数去除特征之间的冗余。为发挥每种特征选择方法的优势，使其优势互补能全面评价特征。本文将 Relief F 评价准则和 EMISE 评价准则融合，提出了结合 Relief F 和 EMISE 算法的改进分离阈值组合式特征优选算法（ReEMISE）。首先，采用 Relief F 算法初选特征子集；然后对初选特征子集通过 ISEaTH 算法求出各特征类间距离与类内距离的比值，并乘于各特征的信息熵；再次，将 ISEaTH 算法计算的特征值赋予 Relief F 算法计算的特征权重值，即 EMISE 特征值与 Relief F 特征权重相乘；最后，计算各特征之间的 Pearson 相关性，通过设置阈值去除特征冗余，得到最优特征子集见图 2。ReEMISE 算法在 Python 3.7 环境下实现。

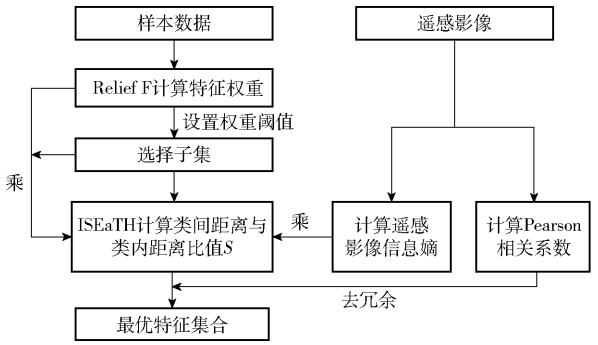

图 2　ReEMISE 特征选择流程图

Fig. 2　Flow chart of ReEMISE feature selection method

2.3　分类方法

随机森林是 BREIMAN[35] 于 2001 年提出的一种非参数分类与回归方法。该算法集成了 Bagging 集成学习理论和随机子空间方法，是一种高效的机器学习算法。由于其能容忍一定的噪声，不需要遵循特定的统计分布，分类精度高，对过度拟合不敏感，能处理高维数据，逐渐成为遥感影像分类应用最广泛的分类器。随机森林算法的基本思想：首选采用 bootstrap 抽样方法，从原始数据中有放回的重复随机抽取 N 个样本子集，大概每个子集的数据占总样本的 2/3，1/3 未被抽到的样本称为袋外数据。然后对 N 个样本子集构建 N 棵决策树，随机抽取 m（$m < M$）个特征作为每棵树分类节点构成的特征集合称为特征的随机子空间，N 棵树产生 N 个分类结果。最后根据 N 个分类结果，采用投票的方式决定待分类样本的类别。

2.4　实验方案

为提高秋季作物分类精度，本文设计 2 种实验方案，见表 1。方案 1 设计 4 组实验对比分析不同特征降维算法的随机森林分类精度以及与支持向量机分类器的对比研究。方案 2 设计 7 组实验对比分析不同特征组合。其研究目的为：① 研究不同特征选择算法对分类精度的影响，分析不同特征的重要性程度。② 通过不同实验对比，研究不同特征组合对分类精度的影响。

表 1　实验方案

Tab. 1　Experimental programs

方案	实验序号	实验内容
不同特征选择算法的分类精度比较	0	Relief F 降维 + 随机森林分类（RF_Relief F）
	1	EMISE 降维 + 随机森林分类（RF_EMISE）
	2	ReEMISE 降维 + 随机森林分类（RF_ReEMISE）
	3	ReEMISE 降维 + 支持向量机分类（SVM_ReEMISE）
不同特征组合的分类精度比较	4	多时相光谱特征
	5	多时相光谱特征 + 传统指数特征
	6	多时相光谱特征 + 红边指数特征
	7	多时相光谱特征 + 纹理特征
	8	多时相光谱特征 + 波段差值特征
	9	多时相光谱特征 + 波段比值特征
	10	特征优选

3　结果与分析

3.1　RF_ReEMISE 分类精度分析

3.1.1　分类结果及精度分析

ReEMISE 特征优选随机森林分类法中，Relief F 特征初选阈值为 100，即选取特征权重排序较大的前 100 个特征组成特征初选子集。本文根据前人研

究成果[32-33,36]以及实验对相关性系数阈值选取0.96。最终得到重要性较高的68个特征见图3,图中,7、8、9分别表示2019年7月2日、2019年8月16日、2019年8月31日影像,r89b11表示2019年8月31日影像b11波段和2019年8月16日影像b11波段的比值,依此类推。在特征数为9个时分类精度最高,总体精度和Kappa系数分别为95.391 8%和0.939 7,见表2。前9个特征重要性得分高,分类精度急剧上升。9个之后特征重要性较低,对分类虽有一定程度的贡献,但特征之间冗余性增强,随机森林分类器性能受影响,分类精度处于波动状态。RF_ReEMIS分类精度混淆矩阵见表2,玉米的生产者精度大于用户精度,造成用户精度低主要是玉米与小宗作物和经济林发生错分,因为小宗作物地块交叉在大片玉米田中,预处理时的重采样使类别边界模糊以及混合像元造成玉米和小宗作物错分,玉米和经济林的光谱相似也造成两者错分。经济林生产者精度较低,主要因为农民在小果树下种植蔬菜以及在道路两旁的景观林下种植大豆造成漏分。小宗作物生产者精度和用户精度都是最低的。主要原因为:① 小宗作物样本少,特征选择和随机森林分类算法的偏向造成。② 常安镇里庄村蔬菜种植地块狭小,种类繁多以及收获时期不同,容易和裸地发生混淆。③ 南董镇的花卉基地地块破碎,实地调研发现花农会在裸地种植小块作物和蔬菜,容易和经济林发生混淆,见图4a。

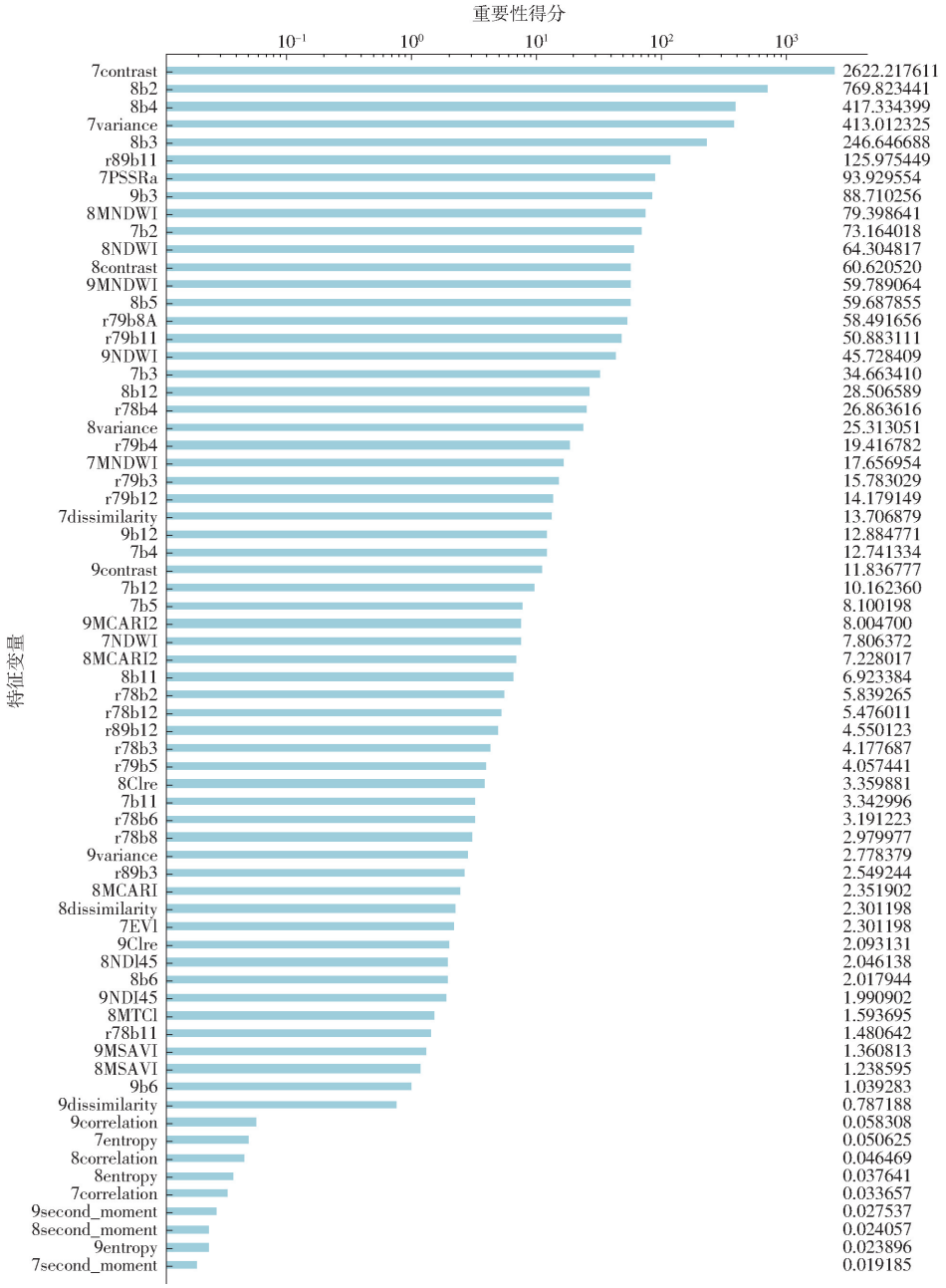

图3 特征重要性

Fig. 3 Distribution of characteristic importance

表 2　ReEMISE 特征选择的随机森林分类精度

Tab. 2　Classification accuracy of RF_ReEMISE

类别	生产者精度/%	用户精度/%
玉米	97.81	95.70
经济林	90.18	98.93
非植被	98.92	98.62
小宗作物	88.40	69.76
大豆	95.58	98.39

3.1.2　选取特征重要性分析

从图 3 中看出，特征重要性差异较大，最高的重要性得分达 2 622.217 611，最低重要性得分只有 0.019 185，对分类几乎不产生影响。重要性得分排序前 9 个特征是 7contrast、8b2、8b4、7variance、8b3、r89b11、7PSSRa、9b3、8MNDWI。纹理特征 7contrast 和 7variance 分别位居第 1 和第 4 位，重要性得分远

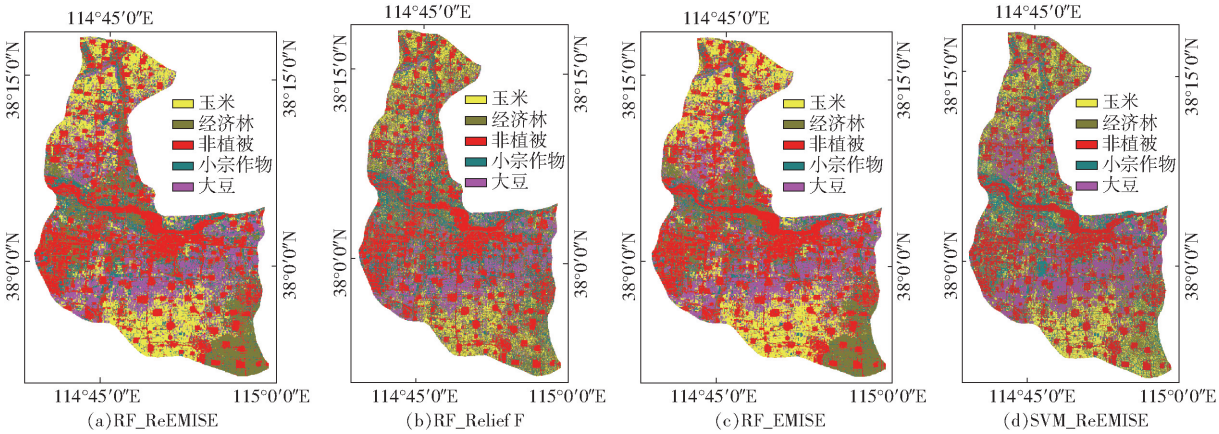

图 4　9 个特征变量分类图

Fig. 4　Classification images of 9 feature variables

高于其他。在前 60 个特征中出现的纹理特征只包括 3 个不同时相的对比度（contrast）、方差（variance）和非相似性（dissimilarity），其都体现了图像的差异性，说明在秋季作物分类中应选择与差异性有关的纹理特征。2019 年 7 月 2 日影像纹理特征重要，主要因为此物候期植被地表覆盖度不同（覆盖度由大到小依次为：经济林、玉米、大豆、非植被）。虽然玉米和大豆都处于苗期，但玉米地表覆盖度大影像亮度较暗，大豆土壤裸露多影像较亮。原始光谱在重要性中占有重要地位，前 9 个特征中 8b2、8b4、8b3、9b3 得分 769.823 441、417.334 399、246.646 688、88.710 256。可知，在 Sentinel－2 的 3 个时相 30 个光谱中可见光波段最重要，其次是短波红外，这与王利民等[37]的研究结果一致，短波红外的加入可以一定程度上提高玉米和大豆的分类精度。红边波段是植被在红光波段和近红外波段之间的一个“爬坡现象”，是 Sentinel－2 卫星除 10 m 分辨率外的又一优势。很多学者对红边波段在植被分类中的重要性进行了研究[38－39]。从特征重要性得分排序看出，红边波段对农作物面积提取有一定帮助，主要体现在红边的 b5 波段，与 MARKUS 等[25]研究结果一致。红边波段排在其次的是 b6 波段，b7 波段被全部剔除。r78b5 和 r78b6 分别排在 40 位和 43 位，不同时相的红边波段比值的重要性反映了红边波段对植被变化比较敏感。近红外波段对农作物提取贡献最小，主要体现在不同时相的短波红外光谱比值，其 r89b11 特征排在第 6 位。水体指数重要性较高主要因为研究区内包括了滹沱河和人工湖，改进的归一化差异水体指数（MNDWI）比归一化水体指数（NDWI）重要性高，说明利用短波红外代替近红外的 MNDWI 提取水体更有效[40]。从特征重要性得分和排序上看，红边植被指数比传统指数中的传统植被指数重要性高，传统植被指数主要体现在能有效降低土壤背景噪声影响的修正土壤调节植被指数（MSAVI）和增强植被指数（EVI）。红边植被指数重要性最高的是冠层叶绿素指数（PSSRa），其次是叶绿素吸收指数（MCARI2）、地面叶绿素指数（MTCI）、红边叶绿素指数（CIre）。红边植被指数的重要性体现在和叶绿素的相关性较大的指数上，说明不同植被叶绿素含量不同，光谱具有差异。利用与叶绿素相关性高的指数可以提高秋季作物分类精度，为农作物种植面积提取提供新思路。分类效果最佳时，9 个特征中不同特征所占的比重见图 5，多时相光谱特征最显著，达到 44.4%，其次是纹理特征，占比 22.2%。

从表 3 中可以看出，8 月中旬是提取农作物的最佳时相，此时主要秋季作物玉米、大豆、谷子和花生等处于茂盛期，光谱信息丰富，同类地物光谱差异小，不易造成“同物异谱”现象。同时，藁城区主要种植的漏天蔬菜菜花处于刚定植时期，生长稀疏，易于光谱识别。其次是 7 月上旬，此时秋季主要作物处于苗期，而果林和景观林长势茂

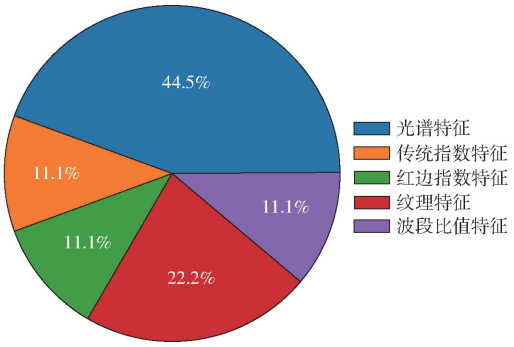

图 5 优选特征子集不同特征变量比重统计

Fig. 5 Proportion statistics of different feature variables of optimized feature subset

盛,易于光谱识别。8 月末和 9 月上旬,此时同种作物播种时间差异光谱反映明显,“同物异谱”现象严重。

表 3 优选特征子集不同时相分布

Tab. 3 Feature subsets preferably distributed in different phases

日期	特征名
2019 - 07 - 02	7contrast、7variance、7PSSRa
2019 - 08 - 16	8b2、8b3、8b4、8MNDWI、r89b11
2019 - 08 - 31	9b3、r89b11

3.2 方案 1 不同特征选择算法及不同分类器精度比较

为比较 ReEMISE 算法特征选择效果,分别与 Relief F 算法和 EMISE 算法的随机森林分类精度比较。同时与支持向量机分类法比较,验证随机森林处理高维小样本数据的性能。4 种不同实验的分类结果见图 4,通过目视解译法与原始图像比较,RF_Relief F 和 SVM_ReEMISE 椒盐现象严重,右下角的梨园存在严重的错分情况,见图 6。为定量评价不同实验的分类精度,实验评价指标包括总体分类精度、错分误差、漏分误差和 Kappa 系数。由表 4 可知,在特征变量 9 个时,ReEMISE 算法达到最高分类精度,其总体精度和 Kappa 系数分别为 95.391 8% 和 0.939 7,相比 EMISE 算法和 Relief F 算法总体分类精度分别提高 1.654 9% 和 24.036 8%,Kappa 系数提高 0.021 6 和 0.309 6。Relief F 算法在特征变量 18 个时达到最高精度,其总体精度和 Kappa 系数分别为 95.340 3% 和 0.939。Relief F 算法在特征变量 45 个时达到最高精度,其总体精度和 Kappa 系数分别为 95.315% 和 0.938 7。由于前 9 个特征重要性得分高,3 种算法分类精度指标都快速上升。ReEMISE 算法在 9 个特征达到最高精度,然后随着特征增加,特征之间出现冗余并且重要性降低,造成精度下降。Relief F 算法特征变量在 9 ~ 18 之间精度指标逐渐上升达到最高,然后处于波动状态。EMISE 算法特征变量 9 个之后精度指标逐渐缓慢上升,45 个时达到最高精度,然后略有下降。从图 7 可以直观看出,利用 ReEMISE 特征优选算法提取农作物,每种类别的错分误差和漏分误差都小于其他算法。ReEMISE 算法错分误差和漏分误差最大的是小宗作物分别为 30.24% 和 11.60%,与 EMISE 算法相比错分误差降低 6 个百分点,漏分误差降低 8.43 个百分点。与 Relief F 算法相比错分误差降低 19.02 个百分点,漏分误差降低 15.18 个百分点。由以上得出,ReEMISE 算法选择子集最优,主要包括 2 个方面:① 最高精度时特征变量数量最少。

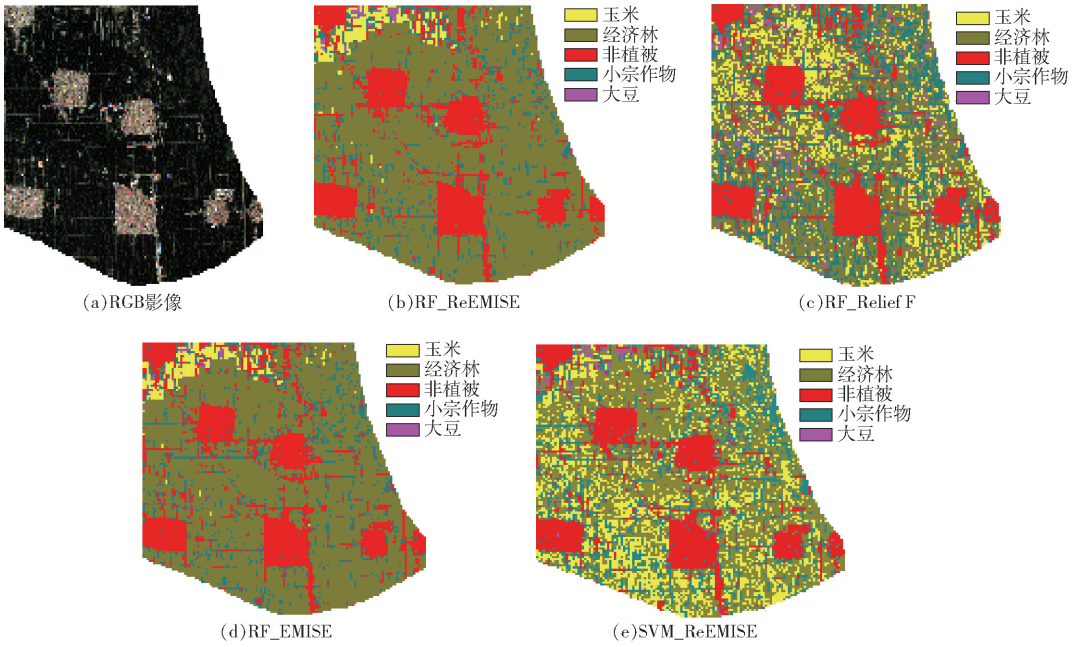

图 6 9 个特征变量局部分类图

Fig. 6 Partial classification images of 9 feature variables

表 4　不同特征选择算法随机森林分类精度比较

Tab. 4　Comparison of classification accuracy of random forest with different feature selections

特征变量	RF_ReEMISE		SVM_ReEMISE		RF_Relief F		RF_EMISE	
个数	总体精度/%	Kappa 系数	总体精度/%	Kappa 系数	总体精度/%	Kappa 系数	总体精度/%	Kappa 系数
3	64. 789 4	0. 544 9	62. 717 2	0. 517 0	62. 576 5	0. 509 3	62. 576 5	0. 509 3
6	75. 652 3	0. 683 5	74. 002 3	0. 661 9	68. 972 1	0. 599 0	75. 618 5	0. 683 0
9	95. 391 8	0. 939 7	78. 431 5	0. 720 6	71. 355 0	0. 630 1	93. 736 3	0. 918 1
12	95. 046 4	0. 935 2	94. 067 0	0. 922 5	89. 368 7	0. 861 8	93. 498 8	0. 915 1
15	94. 883 8	0. 933 0	94. 927 6	0. 933 8	91. 466 4	0. 888 9	94. 204 1	0. 924 2
18	94. 613 4	0. 929 6	94. 799 7	0. 932 1	95. 340 3	0. 939 0	94. 447 1	0. 927 4
21	94. 874 7	0. 932 9	94. 847 2	0. 932 7	95. 218 2	0. 937 4	94. 818 0	0. 932 2
24	94. 748 6	0. 931 3	93. 942 8	0. 920 9	94. 807 0	0. 932 1	94. 734	0. 931 1
27	94. 684 6	0. 930 5	94. 090 8	0. 922 8	95. 048 2	0. 935 2	94. 763 2	0. 931 5
30	95. 024 5	0. 934 9	95. 048 2	0. 935 3	94. 982 5	0. 934 4	94. 849 1	0. 932 7
33	95. 031 8	0. 935 0	95. 135 9	0. 936 5	95. 103 1	0. 935 9	94. 883 8	0. 933 1
36	95. 283 9	0. 938 3	95. 097 6	0. 936 1	94. 993 4	0. 934 5	94. 872 8	0. 933 0
39	95. 220 0	0. 937 5	95. 064 7	0. 935 6	95. 099 4	0. 935 9	94. 757 7	0. 931 4
42	95. 139 6	0. 936 4	95. 103 1	0. 936 2	95. 066 5	0. 935 5	94. 777 8	0. 931 7
45	95. 196 2	0. 937 1	95. 026 3	0. 935 2	94. 856 4	0. 932 7	95. 315 0	0. 938 7
48	95. 079 3	0. 935 7	94. 924 0	0. 933 9	95. 051 9	0. 935 3	95. 024 5	0. 935 0
51	95. 289 4	0. 938 4	95. 123 2	0. 936 4	95. 130 5	0. 936 3	95. 218 2	0. 937 5
54	95. 269 3	0. 938 1	94. 814 4	0. 932 5	94. 909 4	0. 933 4	95. 161 5	0. 936 7
57	95. 360 7	0. 939 3	94. 865 5	0. 933 1	95. 008 0	0. 934 7	94. 944 1	0. 933 9
60	95. 178 0	0. 937 0	94. 962 4	0. 934 4	95. 139 6	0. 936 5	94. 978 8	0. 934 4
63	95. 134 1	0. 936 4	94. 949 6	0. 934 2	95. 073 8	0. 935 6	94. 936 8	0. 933 8
66	95. 254 7	0. 937 9	94. 998 9	0. 936 4	94. 816 2	0. 932 2	95. 139 6	0. 936 5

② 3 种算法不同特征变量数量达到最高精度时 ReEMISE 算法分类精度指标最高。在 ReEMISE 特征选择下，比较随机森林分类器和支持向量机分类器性能。随机森林 9 个特征变量时达到最高精度，比支持向量机总体精度提高 16. 96%，Kappa 系数提高 0. 219 1。支持向量机分类法在特征变量 33 个时达到最高精度，总体精度和 Kappa 系数分别为 95. 135 9% 和 0. 936 5。由表 4 看出，随机森林分类器总体上优于支持向量机分类器。采用 i7 - 9750H CPU，16 GB 内存的测试环境，随机森林分类时，183 个特征全部参与分类用时 610 s，9 个特征分类用时 123 s。支持向量机法在最高精度 33 个特征时分类用时 542 s。ReEMISE 降维算法能在保证精度下提高分类速度。

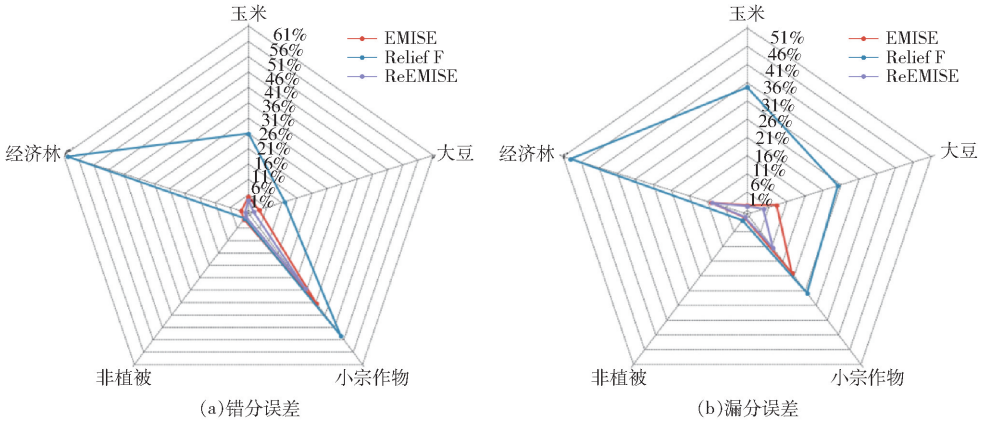

图 7　不同特征选择随机森林分类误差对比

Fig. 7　Comparison of random forest classification errors with different feature selections

3. 3　方案 2 不同特征组合精度分析

为了定量准确评价不同特征组合对农作物提取精度的影响，本文通过在光谱特征基础上分别加入不同特征共设计 7 组实验进行验证，其评价指标主要包括总体精度、Kappa 系数、生产者精度和用户精度，见图 8。多时相光谱特征实验总体精度为 93. 043 4%，Kappa 系数为 0. 908 8。在多时相光谱特征的基础上加入传统指数和红边指数后分类精度

都有所提升,总体精度分别提高 1.502 1、1.571 7 个百分点,Kappa 系数分别提高 0.019 8 和 0.020 7。其中红边指数特征精度最高,说明植被的红边可以显著提高秋季作物分类精度。但加入纹理特征、波段比值特征和波段差值特征后精度都略有下降,其中加入纹理特征后只对非植被精度略有改善,而导致植被类包括玉米、大豆、小宗作物和经济林的生产者和用户精度降低。说明中分辨率 10 m 的遥感影像纹理特征不一定能提高秋季作物分类精度,甚至可能降低精度。加入波段差值和波段比值特征精度下降,说明多时相光谱特征与波段差比特征相关性强,对分类贡献小。这与图 3 特征的重要性排序结果一致,从 183 个特征中选择重要性较高的 68 个特征,不同时相的波段差值特征全部被剔除。加入传统指数,除了非植被类别用户精度略微降低外,其他类别用户精度和生产者精度都有提升,尤其对经济林、小宗作物和大豆的生产者精度分别提高 2.89、2.26、2.61 个百分点。说明传统指数的加入有助于提高经济林、小宗作物和大豆的精度。加入红边指数,经济林的生产者精度提高最显著,由 83.72% 提高到 90.94%。加入红边指数对提高大豆和玉米的精度也有帮助,但小宗作物的生产者精度下降 1.4 个百分点。波段差值导致各类别用户精度和生产者精度不增反降。波段比值除大豆生产者精度和小宗作物用户精度下降外,其他都有不同程度提高。说明不同时相波段比值特征比波段差值特征对提高秋季作物分类精度更有帮助。ReEMISE 算法特征优选结果与多时相光谱相比总体精度提高 2.348 4 个百分点,Kappa 系数提高 0.030 9。特征优选中各类别的生产者精度和用户精度都比其他特征组合的精度高,说明多种特征组合可以有效提高秋季作物分类精度。

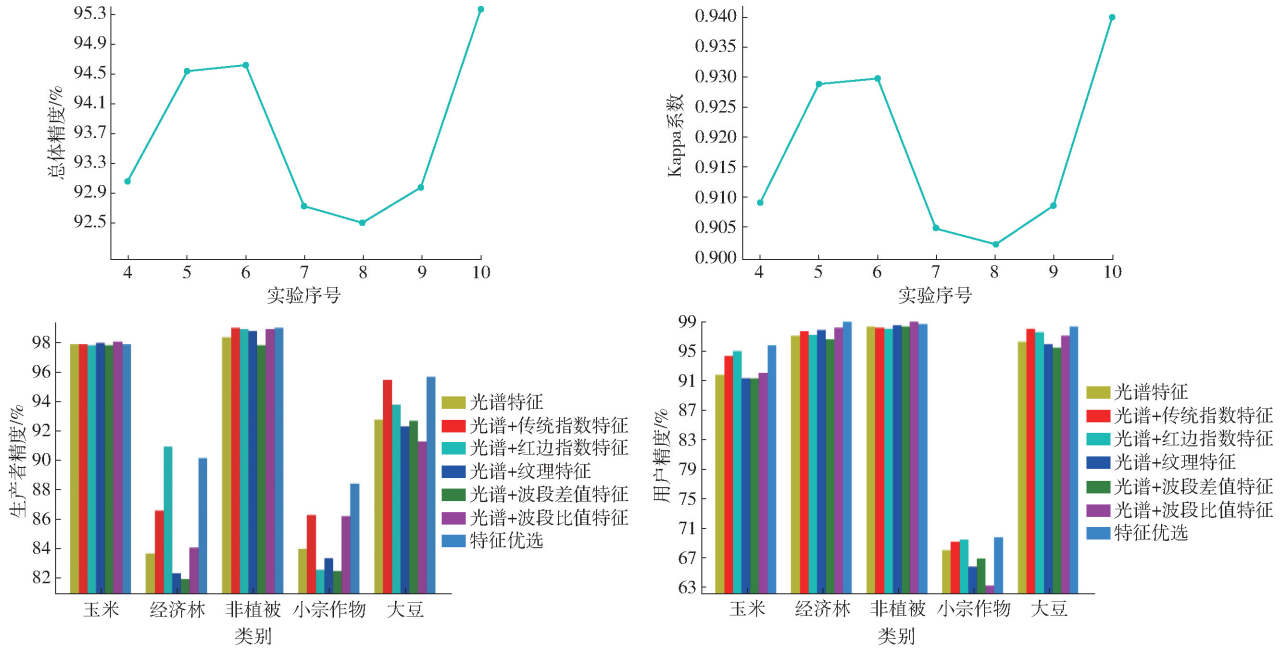

图 8 不同特征组合精度对比

Fig. 8 Comparison of accuracy of different feature combinations

4 结论

(1)在随机森林分类前,利用 ReEMISE 降低特征维度可以有效提高运算速度和分类精度。相比 Relief F 和 EMISE 降维算法,ReEMISE 在保证分类精度的前提下,可大幅降低特征维度,说明 ReEMISE 优选的特征子集数据冗余小、特征重要性高,能很好地平衡特征维度和分类精度,与支持向量机分类器相比,随机森林分类精度更高。

(2)在多时相光谱特征基础上,通过单独引入不同特征研究了不同特征对秋季作物分类精度的影响。传统指数特征和红边指数特征对分类精度产生积极影响,分别加入纹理特征、不同时相波段比值特征和不同时相波段差值特征后分类精度有所下降。全部特征参与的特征优选组合生产者精度和用户精度都比其他组合高,综合多特征是提高秋季作物分类精度的关键。

(3)特征重要性分析表明,原始光谱信息的 8b2(2019 年 8 月 16 日影像蓝光)、8b4(2019 年 8 月 16 日影像红光)、8b3(2019 年 8 月 16 日影像绿光)、9b3(2019 年 8 月 31 日影像绿光)最重要,其次是纹理信息的 7contrast(2019 年 7 月 2 日影像对比度)、7variance(2019 年 7 月 2 日影像方差),证明了能表现差异性的纹理特征在农作物分类中的重要性。不同时相波段的差值特征全部被舍弃。

参 考 文 献

[1] 黄健熙, 侯裔焯, 苏伟,等. 基于 GF－1 WFV 数据的玉米与大豆种植面积提取方法[J]. 农业工程学报, 2017, 33(7): 164－170.
HUANG Jianxi, HOU Yuzhuo, SU Wei, et al. Mapping corn and soybean cropped area with GF－1 WFV data[J]. Transactions of the CSAE, 2017, 33(7): 164－170. (in Chinese)

[2] 刘佳, 王利民, 杨福刚, 等. 基于 HJ 时间序列数据的农作物种植面积估算[J]. 农业工程学报,2015,31(3):199－206.
LIU Jia, WANG Limin, YANG Fugang, et al. Remote sensing estimation of crop planting area based on HJ time-series images[J]. Transactions of the CSAE,2015,31(3):199－206. (in Chinese)

[3] 赵子娟, 刘东, 杭中桥. 物遥感识别方法研究现状及展望[J]. 江苏农业科学,2019,47(16):45－51.
ZHAO Zijuan, LIU Dong, HANG Zhongqiao. Research status and prospect of object remote sensing identification method[J]. Jiangsu Agricultural Sciences, 2019,47(16):45－51. (in Chinese)

[4] 张磊, 宫兆宁, 王启为, 等. Sentinel－2 影像多特征优选的黄河三角洲湿地信息提取[J]. 遥感学报,2019,23(2):313－326.
ZHANG Lei, GONG Zhaoning, WANG Qiwei, et al. Wetland mapping of Yellow River Delta wetlands based on multi-feature optimization of Sentinel－2 images[J]. Journal of Remote Sensing, 2019,23(2): 313－326. (in Chinese)

[5] 刘扬, 付征叶, 郑逢斌. 高分辨率遥感影像目标分类与识别研究进展[J]. 地球信息科学学报,2015,17(9):1080－1091.
LIU Yang, FU Zhengye, ZHENG Fengbin. Review on high resolution remote sensing image classification and recognition[J]. Journal of Geo-information Science,2015,17(9): 1080－1091. (in Chinese)

[6] 王利军, 郭燕, 贺佳, 等. 基于决策树和 SVM 的 Sentinel－2A 影像作物提取方法[J/OL]. 农业机械学报, 2018,49(9): 146－153.
WANG Lijun, GUO Yan, HE Jia, et al. Classification method by fusion of decision tree and SVM based on Sentinel－2A image[J/OL]. Transactions of the Chinese Society for Agricultural Machinery,2018,49(9):146－153. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no = 20180917&flag = 1. DOI: 10.6041/j.issn.1000-1298.2018.09.017. (in Chinese)

[7] 郭伟, 赵春江, 顾晓鹤, 等. 乡镇尺度的玉米种植面积遥感监测[J]. 农业工程学报,2011,27(9):69－74.
GUO Wei, ZHAO Chunjiang, GU Xiaohe, et al. Remote sensing monitoring of maize planting area at town level[J]. Transactions of the CSAE,2011,27(9):69－74. (in Chinese)

[8] 马丽. 多源信息复合的 SVM 混合地块分解法提取玉米种植面积[D]. 西安:西安科技大学,2009.
MA Li. Extracting corn planting area by multi-source data with SVM mixed-field decomposed method[D]. Xi'an: Xi'an University of Technology,2009. (in Chinese)

[9] 邓曾, 李丹, 柯樱海, 等. 基于改进 SVM 算法的高分辨率遥感影像分类[J]. 国土资源遥感,2016,28(3):12－18.
DENG Zeng, LI Dan, KE Yinghai, et al. An improved SVM algorithm for high spatial resolution remote sensing image classification[J]. Remote Sensing for Land and Resources,2016,28(3):12－18. (in Chinese)

[10] 李龙, 李旭青, 吴伶, 等. 基于决策树和神经网络的农作物分类研究——以廊坊市为例[J]. 红外,2019, 40(3):24－31.
LI Long, LI Xuqing, WU Ling, et al. Crop classification research based on decision tree and neural network—take Langfang City as an example[J]. Infrared,2019,40(3):24－31. (in Chinese)

[11] 余强毅, 吴文斌, 陈羊阳, 等. 农作物空间格局变化模拟模型的 MATLAB 实现及应用[J]. 农业工程学报, 2014, 30(12):105－114.
YU Qiangyi, WU Wenbin, CHEN Yangyang, et al. Model application of an agent-based model for simulating crop pattern dynamics at regional scale based on MATLAB[J]. Transactions of the CSAE,2014,30(12):105－114. (in Chinese)

[12] 何昭欣, 张淼, 吴炳方, 等. Google Earth Engine 支持下的江苏省夏收作物遥感提取[J]. 地球信息科学学报, 2019, 21(5):752－766.
HE Zhaoxin, ZHANG Miao, WU Bingfang, et al. Extraction of summer crop in Jiangsu based on Google Earth Engine[J]. Journal of Geo-information Science, 2019, 21(5): 752－766. (in Chinese)

[13] 黄双燕, 杨辽, 陈曦, 等. 机器学习法的干旱区典型农作物分类[J]. 光谱学与光谱分析, 2018, 38(10):3169－3176.
HUANG Shuangyan, YANG Liao, CHEN Xi, et al. Study of typical arid crops classification based on machine learning[J]. Spectroscopy and Spectral Analysis, 2018, 38(10): 3169－3176. (in Chinese)

[14] 陈启浩. 面向对象的多源遥感数据分类技术研究与实现[D]. 武汉:中国地质大学,2007.
CHEN Qihao. Research and implementation of object-oriented multi-source remote sensing data classification technology[D]. Wuhan: China University of Geosciences,2007. (in Chinese)

[15] HYUN O K, JONG M Y. Effect of red-edge and texture features for object-based paddy rice crop classification using RapidEye multi-spectral satellite image data[J]. International Journal of Remote Sensing, 2014,35(19):7046－7068.

[16] 杜保佳, 张晶, 王宗明, 等. 应用 Sentinel－2A NDVI 时间序列和面向对象决策树方法的农作物分类[J]. 地球信息科学学报,2019,21(5):740－751.
DU Baojia, ZHANG Jing, WANG Zongming, et al. Crop mapping based on Sentinel－2A NDVI time series using object-oriented classification and decision tree model[J]. Journal of Geo-information Science, 2019, 21(5): 740－751. (in Chinese)

[17] 程希萌, 沈占锋, 邢廷炎, 等. 基于 mRMR 特征优选算法的多光谱遥感影像分类效率精度分析[J]. 地球信息科学学报, 2016,18(6): 815－823.
CHENG Ximeng, SHEN Zhanfeng, XING Tingyan, et al. Efficiency and accuracy analysis of multispectral image classification based on mRMR feature selection method[J]. Journal of Geo-information Science, 2016, 18(6): 815－823. (in Chinese)

[18] 宋羚. 基于多目标遗传算法和 SVM 的特征选择方法[D]. 武汉:华中科技大学,2007.
SONG Ling. A feature selection method based on multi-objective genetic algorithm and support vector machines[D]. Wuhan: Huazhong University of Science and Technology,2007. (in Chinese)

[19] 刘家福, 李林峰, 任春颖. 基于特征优选的随机森林模型的黄河口滨海湿地信息提取研究[J]. 湿地科学,2018, 16(2):97－105.
LIU Jiafu, LI Linfeng, REN Chunying. Information extraction of coastal wetlands in Yellow River Delta by optimal feature-

based random forest model[J]. Wetland Science,2018,16(2):97 - 105. (in Chinese)
[20] 刘舒, 姜琦刚, 马玥, 等. 基于多目标遗传随机森林特征选择的面向对象湿地分类[J/OL]. 农业机械学报,2017,48(1):119 - 127.
LIU Shu, JIANG Qigang, MA Yue, et al. Object-oriented wetland classification based on hybrid feature selection method combining with Relief F, multi-objective genetic algorithm and random forest[J/OL]. Transactions of the Chinese Society for Agricultural Machinery,2017,48(1):119 - 127. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no = 20170116&flag = 1. DOI: 10.6041/j.issn.1000-1298.2017.01.016. (in Chinese)
[21] 何云, 黄翀, 李贺, 等. 基于 Sentinel - 2A 影像特征优选的随机森林土地覆盖分类[J]. 资源科学,2019,41(5): 992 - 1001.
HE Yun, HUANG Chong, LI He, et al. Land-cover classification of random forest based on Sentinel - 2A image feature optimization[J]. Resources Science, 2019, 41(5): 992 - 1001. (in Chinese)
[22] 苏伟, 张明政, 蒋坤萍, 等. Sentinel - 2 卫星影像的大气校正方法[J]. 光学学报,2018,38(1):314 - 323.
SU Wei, ZHANG Mingzheng, JIANG Kunping, et al. Atmospheric correction method for Sentinel - 2 satellite imagery[J]. Acta Optica Sinica, 2018, 38(1): 314 - 323. (in Chinese)
[23] 吴静, 吕玉娜, 李纯斌, 等. 基于多时相 Sentinel - 2A 的县域农作物分类[J/OL]. 农业机械学报,2019,50(9):194 - 200.
WU Jing, LÜ Yu'na, LI Chunbin, et al. Fine classification of county crops based on multi-temporal images of Sentinel - 2A [J/OL]. Transactions of the Chinese Society for Agricultural Machinery,2019,50(9):194 - 200. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no = 20190922&flag = 1. DOI: 10.6041/j.issn.1000-1298.2019.09.022. (in Chinese)
[24] 刘佳, 王利民, 滕飞, 等. RapidEye 卫星红边波段对农作物面积提取精度的影响[J]. 农业工程学报,2016,32(13):140 - 148.
LIU Jia, WANG Limin, TENG Fei, et al. Impact of red-edge waveband of RapidEye satellite on estimation accuracy of crop planting area [J]. Transactions of the CSAE,2016, 32(13): 140 - 148. (in Chinese)
[25] MARKUS I, FRANCESCO V, CLEMENT A. First experience with Sentinel - 2 data for crop and tree species classifications in central Europe[J]. Remote Sensing, 2016, 8(3):166 - 192.
[26] 王娜, 李强子, 杜鑫, 等. 单变量特征选择的苏北地区主要农作物遥感识别[J]. 遥感学报, 2017,21(4):519 - 530.
WANG Na, LI Qiangzi, DU Xin, et al. Identification of main crops based on the univariate feature selection in Subei[J]. Journal of Remote Sensing,2017,21(4): 519 - 530. (in Chinese)
[27] 张卫春, 刘洪斌, 武伟. 基于随机森林和 Sentinel - 2 影像数据的低山丘陵区土地利用分类——以重庆市江津区李市镇为例[J]. 长江流域资源与环境, 2019,28(6):1334 - 1343.
ZHANG Weichun, LIU Hongbin, WU Wei. Classification of land use in low mountain and hilly area based on random forest and Sentinel - 2 satellite data: a case study of Lishi Town, Jiangjin, Chongqing [J]. Resources and Environment in the Yangtze Basin, 2019, 28(6): 1334 - 1343. (in Chinese)
[28] 方灿莹, 王琳, 徐涵秋. 不同植被红边指数在城市草地健康判别中的对比研究[J]. 地球信息科学学报,2017,19(10): 1382 - 1392.
FANG Canying, WANG Lin, XU Hanqiu. A comparative study of different red edge indices for remote sensing detection of urban grassland health status[J]. Journal of Geo-information Science,2017, 19(10):1382 - 1392. (in Chinese)
[29] KIRA K, RENDELL L. The feature selection problem: traditional methods and a new algorithm[C]//Proceedings of the Ninth National Conference on Artificial Intelligence. New Orleans: AAAI Press,1992:129 - 134.
[30] KONONENKO I. Estimation attributes: analysis and extensions of RELIEF[C]//Proceedings of the 1994 European Conference on Machine Learning,1994: 171 - 182.
[31] NUSSBAUM S, NIEMEYER I, CANTY M J. SEaTH - a new tool for automated feature extraction in the context of object-based image analysis for renote sensing[C]//The lst International Conference on Object-based Image Analysis, Salzhourg, Austria, 2006.
[32] 余晓敏, 湛飞并, 廖明生, 等. 利用改进 SEaTH 算法的面向对象分类特征选择方法[J]. 武汉大学学报(信息科学版), 2012, 37(8): 921 - 924.
YU Xiaomin, ZHAN Feibing, LIAO Mingsheng, et al. Object-oriented feature selection algorithms based on improved SEaTH algorithms[J]. Geomatics and Information Science of Wuhan University, 2012, 37(8): 921 - 924. (in Chinese)
[33] 杨惠雯, 方俊永, 赵冬. 基于改进分离阈值法的农作物遥感精细分类特征选择[J]. 应用科学学报, 2019, 37(1):51 - 63.
YANG Huiwen, FANG Junyong, ZHAO Dong. Feature selection in precise crop classification using remote sensing data based on improved SEaTH algorithm[J]. Journal of Applied Sciences, 2019, 37(1): 51 - 63. (in Chinese)
[34] 吴浩苗,尹中航,孙富春. Relief 算法在笔迹识别中的应用[J]. 计算机应用,2006,26(1): 174 - 176.
WU Haomiao, YIN Zhonghang, SUN Fuchun. The application of Relief algorithm in handwriting recognition[J]. Computer Science and Application,2006,26(1):174 - 176. (in Chinese)
[35] BREIMAN L. Random forests[J]. Machine Learning,2001,45(1):5 - 32.
[36] 杨宁. 高分辨率影像面向对象分类特征选择方法研究[D]. 西安:西安科技大学,2012.
YANG Ning. Research on feature selection method for object-oriented classification of high resolution image[D]. Xi'an: Xi'an University of Technology,2012. (in Chinese)
[37] 王利民, 刘佳, 杨玲波, 等. 短波红外波段对玉米大豆种植面积识别精度的影响[J]. 农业工程学报,2016,32(19): 169 - 178.
WANG Limin, LIU Jia, YANG Lingbo, et al. Impact of short infrared wave band on identification accuracy of corn and soybean area[J]. Transactions of the CSAE, 2016, 32(19): 169 - 178. (in Chinese)
[38] GERALD F, KANGBENI D, IDRISS S, et al. Landsat - 8 *vs* Sentinel - 2: examining the added value of sentinel - 2's red-edge bands to land-use and land-cover mapping in Burkina Faso[J]. GIScience & Remote Sensing,2018,55(3):331 - 354.
[39] RADOUX J, CHOMÉ G, JACQUES D C, et al. Sentinel - 2's potential for sub-pixel landscape feature detection[J]. Remote Sensing,2016,8(6):488 - 517.
[40] QI J, CHEHBOUNI A, HUETE A R, et al. A modified soil adjusted vegetation index[J]. Remote Sensing Environment, 2015,48(2):119 - 126.