

多尺度分割与特征优选下的盐碱地提取

朱 丽¹, 国巧真^{1*}, 吴正鹏², 吴欢欢¹, 何云海¹

1. 天津城建大学 地质与测绘学院, 天津 300384

2. 天津市测绘院有限公司, 天津 300381

摘 要：土地盐渍化作为一种土壤灾害，严重制约着社会经济与农业的发展。对盐碱地进行实时监测，可为盐碱地的评价改良提供科学依据。由于盐碱地的信息复杂、提取精度不高，因此本文以高分六号（GF-6）卫星遥感影像为数据源，采用分形网络演化算法（fractal net evolution approach, FNEA）进行影像对象的多尺度分割，从面向对象的角度减少高分影像分类结果中的椒盐噪声问题，通过计算图像对象的局部方差和变化率来确定适宜的盐碱地分割尺度。利用基于特征选择的相关性算法（correlations-based feature selection, CFS）与 Relief F 算法分别对由光谱、纹理、形状、遥感指数构成的初始特征空间进行特征优选，精简特征子集，解决特征数量冗余问题，以此来优化随机森林对盐碱地提取精度。结果表明：CFS 约简后的特征子集更小，精度更高，说明在盐碱地提取过程中，筛选特征数目能够减小冗余数据对提取精度的影响。CFS 优化后的随机森林对盐碱地的提取效果较好，该方法总体分类精度达到 83.7%。

关键词：GF-6；盐碱地；面向对象；特征选择；随机森林

Extraction of saline-alkali land based on multi-scale segmentation and feature optimization

ZHU Li¹, GUO Qiaozhen^{1*}, WU Zhengpeng², WU Huanhuan¹, HE Yunhai¹

1. School of Geology and Geomatics, Tianjin Chengjian University, Tianjin 300384, China

2. Tianjin Institute of Surveying and Mapping Co., Ltd., Tianjin 300381, China

Abstract: Background, aim, and scope As a kind of soil disaster, soil salinization restricts the development of social economy and agriculture seriously. Real-time monitoring of saline-alkali land can provide scientific basis for evaluation and improvement of saline-alkali land. This paper mainly discussed the effect of object-oriented feature optimization algorithm applied to saline-alkali land. Taking Binhai New Area of Tianjin as the research area and the saline-alkali land as the research object, an optimization model was established to extract saline-alkali land. **Materials and methods** Taking GF-6 satellite remote sensing image as the data source, from the perspective of object-oriented, FNEA was used to segment the image object in multi-scales. Based on the statistics of local variances and change rates under different scales, the appropriate scale for saline-alkali land

收稿日期：2021-11-09；录用日期：2022-05-18；网络出版：2022-05-31

Received Date: 2021-11-09; Accepted Date: 2022-05-18; Online first: 2022-05-31

基金项目：天津市自然科学基金项目（18JCYBJC90900）；国家自然科学基金项目（41971310）

Foundation Item: Natural Science Foundation of Tianjin (18JCYBJC90900); National Natural Science Foundation of China (41971310)

通信作者：国巧真, E-mail: gqiaozhen@tcu.edu.cn

Corresponding Author: GUO Qiaozhen, E-mail: gqiaozhen@tcu.edu.cn

引用格式：朱 丽, 国巧真, 吴正鹏, 等. 2022. 多尺度分割与特征优选下的盐碱地提取 [J]. 地球环境学报, 13(6): 714–723.

Citation: Zhu L, Guo Q Z, Wu Z P, et al. 2022. Extraction of saline-alkali land based on multi-scale segmentation and feature optimization [J]. Journal of Earth Environment, 13(6): 714–723.

recognition was selected. On this basis, the initial feature space was constructed from four perspectives: spectral feature, texture feature, shape feature and remote sensing index feature. Two algorithms, CFS and Relief F algorithm, was used to optimize the initial feature space respectively. The obtained feature subset was used to optimize the extraction effect of random forest algorithm on salinized land, and the two optimization results were compared and discussed. **Results** (1) FNEA algorithm was used for multi-scale segmentation, and the appropriate segmentation scale of saline-alkali land in the study area was 123. (2) The initial feature space was optimized by CFS algorithm and Relief F, and the number of features was reduced to 40 and 17 respectively. (3) The overall classification accuracy of random forest extraction of saline alkali land was 76.3%, and that of random forest optimized by Relief F algorithm was 77.4%. The overall classification accuracy of CFS-optimized random forest in salt and alkali extraction was 83.7%. **Discussion** These results indicated that the CFS and Relief F, as two classical data filtering algorithm, for feature selection, can improve accuracy of random forest model in saline-alkali land, and can make the model improved to a certain extent. The overlap rate feature subsets optimized by the two algorithms was as high as 82%, indicating that the two kinds of algorithm of important characteristics were good search results. In addition, the features filtered out by CFS algorithm were almost twice as many as those filtered out by Relief F, but CFS optimized random forest had a better extraction effect on saline-alkali land, which indicated that the number of features was not positively correlated with the final classification accuracy. **Conclusions** Compared with Relief F algorithm, the random forest model optimized by CFS algorithm had a better recognition effect on saline-alkali land extraction, and the overall accuracy was 83.7%, which was 7.4% higher than that before optimization. CFS algorithm reduced the features to 17 and filtered 81.7% of the features, which solved the problem of data redundancy to a certain extent and improved the quality of data subset and the operating efficiency of stochastic forest algorithm. **Recommendations and perspectives** Firstly, The research ideas proposed in this paper can solve the problem of machine learning capability degradation caused by high-dimensional data redundancy, and can be applied to the optimization of other algorithm models and the recognition and extraction of ground classes. Secondly, in view of the different characteristics of saline-alkali land in different seasons, the influence of seasonal change on saline-alkali land should be considered in the future research.

Key words: GF-6; saline-alkali soil; object-oriented; feature selection; random forest

盐碱地是盐土和碱土的总称,其形成的根本原因是土壤中水盐失衡导致盐分在土壤表面移动与积累(翁永玲和宫鹏,2006)。盐碱地的碱性性质会抑制甚至危害作物的生长发育,因此盐碱地的存在会造成粮食减产、土壤退化等问题,严重制约经济和生态环境的发展。盐碱地作为潜在的土地资源,对其进行监测与改良,对于缓解土地资源紧张、挖掘农业发展潜力具有重要意义。

传统的盐碱地监测主要采用实地土壤调查取样的方法,通过分析土壤各组分含量来验证土壤类型,其精度较高,但在大范围区域监测中,该方法需要消耗大量的社会资源且及时性不强,实现实时动态监测存在一定的困难。随着空间信息技术的不断发展,国内外已经开始广泛利用影像数据如Landsat、QuickBird、SPOT、IKONOS、

GF-1 等对土地的盐渍化信息进行提取,以此提高土地盐渍化的监测效率(Elnaggar and Noller, 2009; Ivits et al, 2013; Allbed et al, 2014; Sidike et al, 2014; 牛增懿等, 2016)。由于高空间分辨率遥感影像包含了更丰富的地物信息,因此选择合适的图像分割算法能够提高目标识别的精确度和稳健性(高仁强等, 2020)。随着影像分辨率的不断提高和影像特征的不断增加,传统的图像分割方法如阈值分割、区域合成、边缘检测等方法也在不断发展,分形网络演化算法(FNEA)作为一种基于区域合成的多尺度分割算法,能综合考虑地物之间的光谱、纹理信息等特征差异,目前已被广泛运用于图像分割中。

从 20 世纪 80 年代至今,国内外针对盐碱地的提取方法研究也在不断发展,决策树、支持向量

机、神经网络、随机森林等机器学习方法也普遍用于盐碱地信息的提取和反演（李晋等，2014；姜红等，2017；徐存东等，2018；Jiang et al, 2019；Wang et al, 2019；杨练兵等，2021），随机森林作为一种并行的集成学习算法，在决策树的基础上进行集成，突破了单分类器的性能提升瓶颈，各个树之间的独立运行让它可以在高维数据上实现并行处理，但其在高维数据上性能仍有提升空间（王奕森和夏树涛，2018）。Rodríguez et al（2006）通过主成分变换（PCA）对随机森林的特征进行降维，但使用该方法时只保留了主成分值较大的值，一些主成分值小但相关性强的特征会被过滤掉。还有学者通过构建特征子空间来优化随机森林，主要思想都是通过分析特征的信息量和相互关系来构建算法（Amaratunga et al, 2008；Ye et al, 2013）。盐碱地信息由于受到季节变化、土壤湿度等影响，利用机器学习需要充分考虑盐碱地的信息复杂性，需结合盐碱地的形状、纹理等特征，但当特征维数过大，会造成数据冗余和无关特征的增加，反而会导致机器学习能力下降，也会使得分类精度降低（Cui et al, 2020），因此在机器学习前需进行特征优选，进而提高机器学习的分类性能。目前特征优选在干旱区精细植被分类、沙化土地识别、湿地分类等方向皆有所应用，而在盐碱地提取方向，仍需要进一步的研究（李长龙等，2015；张磊等，2019；张文博等，2021）。

本文基于 GF-6 多光谱影像数据，利用分形网络演化算法（FNEA）对影像对象进行多尺度分割，针对多维数和数据不平衡问题，使用数据挖掘中的经典算法——CFS 算法与 Relief F 算法进行数据降维，精简特征子集，再利用这两种特征优选算法对面向对象的随机森林算法进行优化。

1 研究区概况及数据源

1.1 研究区概况

研究区位于天津市滨海新区东北部，位置范围北纬 38°40′—39°00′，东经 117°20′—118°00′。天津市北面区域紧邻燕山山脉，山区南部与华北平原地区相连，自此至东南区域地势总体上平缓，海拔在 8 m 以下，一般处于 3—5 m。滨海新区濒临渤海，处于海陆交接处，多年以来受海水浸渍土壤，在温带季风气候影响下，年降水量在

500 mm 左右，而年蒸发量达到降水量的 4 倍（杨晓潇等，2019），因此导致该地区土壤的浅层不断积累盐分，从而形成盐渍化土壤。其次，工业化的迅速发展、农业耕作中不合理的灌溉方式也在加重该区域的土壤盐渍化程度。由于土壤含盐量高，肥力低，区域正常植被生长受到抑制，多生长碱蓬（*Suaeda glauca*）、怪柳（*Tamarix chinensis*）等盐生植物。本文选择研究区域位于滨海新区北大港水库附近，研究区大小为 1500×1500 个像元，研究区内主要包含的土地利用类型为裸露盐碱地、盐生植被、建筑区、道路、水体等地物，土地盐渍化程度不均匀，土壤环境复杂。

1.2 数据源

研究区采用的数据源为高分六号 PMS 多光谱波段数据，主要参数如表 1 所示。研究区遥感影像获取时间为 2019 年 9 月 30 日，处于夏季早期，植被生长旺盛，日照强烈，降水少且蒸发量大，在盐碱地形成过程中正处于脱盐末期，较适合对其进行提取。

表 1 高分六号主要参数 Tab.1 Main parameters of GF-6	
光谱范围 Spectral range	蓝谱段 B1: 0.45—0.52 μm Blue band B1: 0.45—0.52 μm
	绿谱段 B2: 0.52—0.60 μm Green band B2: 0.52—0.60 μm
	红谱段 B3: 0.63—0.69 μm Red band B3: 0.63—0.69 μm
	近红外谱段 B4: 0.76—0.90 μm Near infrared band B4: 0.76—0.90 μm
	分辨率 Resolution
	多光谱: 8 m Multispectral: 8 m
幅宽 Image width	90 km

2 研究方法

2.1 多尺度分割

高空间分辨率遥感影像的优势在于展现地物丰富的形态特征，对于小目标的识别能力也更强，利用传统基于像元的分类方法往往会造成空间数据的冗余，产生椒盐图像，进而降低分类效果。本文选取面向对象方法，其基本思想是通过综合分析不同对象在特征和属性上的差异，将具有相同特质的像元归为一个研究对象。面向对象的分类方法过程主要包含影像分割、影像对象构建、分类规则的建立、信息提取（陈云浩等，2006）。

影像分割是面向对象分类方法中关键的过程之一,其分割尺度的确定对影像分类精度有直接影响。本文利用一种基于区域生长的多尺度分割算法——分形网络演化算法(FNEA)进行图像分割,该算法通过将影像对象间的平均异质性(average heterogeneity)最小化,并将其各自的同质性(homogeneity)最大化,综合考虑影像的光谱和空间纹理信息,基于成对区域合并技术进行自下而上的影像对象合并(Benz et al, 2004)。该算法中异质性由光谱异质性与形状异质性共同决定,其中形状异质性又由紧致度与光滑度两部分组成。FNEA算法(张萌, 2019)如下:

$$H = \omega \times H_{\text{color}} + (1 - \omega) \times H_{\text{shape}} \quad (1)$$

$$H_{\text{color}} = \sum_n (\omega_n \times \sigma_n) \quad (2)$$

$$H_{\text{shape}} = \omega_{\text{smooth}} \times H_{\text{smooth}} + \omega_{\text{compact}} \times H_{\text{compact}} \quad (3)$$

式中: H 为异质性值; ω 为光谱异质性值的权重; H_{color} 为光谱异质性值; H_{shape} 为形状异质性值; ω_n 为某一波段权重; σ_n 为某一波段的像元标准差; H_{smooth} 为光滑度; H_{compact} 为紧致度。

不同的地物类型,由于各自的属性特征,在不同分割尺度上有不同的分割效果,因此本文利用ESP(estimation of scale parameter)尺度评价工具进行盐碱地的最优尺度选择,多尺度分割中的尺度参数需要进行多次人为调试,而ESP则通过计算不同分割尺度下影像对象的局部变化(local variance, LV),通过变化率(rate of change, ROC)峰值来确定适宜的分割尺度,来消除人工调试的主观因素影响(Drăguț et al, 2010)。由于影像中存在多种地物,通过计算得到的ROC峰值一般也不止1个,针对出现的若干分割尺度,需要进行试验确定地物对应的分割尺度。

2.2 特征选择

Relief F算法是Kononenko et al(1997)在只适用于处理二分类问题的Relief算法上进行改进的支持多分类的数据分析方法。作为数据挖掘中经典的Filter算法,其基本思想是对每一个特征进行评价,根据每一个特征与已定义样本类别的相关性,赋予特征权重,每个特征的权重是通过在样本集 D 中随机选择一个样本 S ,计算样本 S 特征值与同类的其他特征值的 k 个最近邻距离与不同类样本的 k 个最近邻距离,通过循环迭代 M 次,类别相关性高的特征将会赋予高的权重。权重计算公式(何牧宇和周晖, 2019)如下:

$$\omega_A = \omega_A - \sum_{i=1}^k \frac{\text{diff}(A, S, H_i)}{k \times M} + \sum_{C \notin \text{class}(S)} \left[\sum_{i=1}^k \frac{\text{diff}(A, S, M_i)}{k \times M} \times \frac{P_C}{1 - P_{\text{class}(S)}} \right] \quad (4)$$

式中: ω_A 为特征 A 的权重; H_i 为与 S 同类的最近邻样本; M_i 为与 S 不同类的最近邻样本; P_C 为类别为 C 的概率; $\text{class}(S)$ 为样本 S 所属的同类别样本子集; $\text{diff}(A, S, M_i)$ 为样本 S_1 与 S_2 在特征 A 上的差,当特征为数值变量时,将数据归一化至 $[0, 1]$ 。

CFS算法是一种关联性的Filter算法,通过计算特征与类别、特征与特征之间的相关性进行评估,从而实现数据清洗(Li et al, 2011)。CFS算法首先针对初始特征空间,采用前向选择或后向选择进行特征子空间的搜索,构建特征子空间 T ,基于启发式估计方法对特征子空间内特征与特征、特征与类别间的相关性进行评估,其相关性强弱利用皮尔逊相关系数进行计算,去除特征与类别间相关性低的特征以及特征与特征相关性过高的特征。启发式评估公式(孙宁青, 2010)如下:

$$M_T = \alpha \times \bar{r}_{\text{cf}} \div \sqrt{\alpha + \alpha \times (\alpha - 1) \times \bar{r}_{\text{ff}}} \quad (5)$$

式中: M_T 为特征子集 T 的评估值; $\bar{r}_{\text{cf}}$ 为类别与特征的平均相关性; $\bar{r}_{\text{ff}}$ 为特征与特征间的相关性; α 表示特征子集包含的特征个数。

2.3 影像分类算法

随机森林(random forest, RF)算法作为集成决策树的机器学习方法,通过在训练集中随机抽取样本且放回的方法,进行每一个决策树的无剪枝生长,来消除决策树受训练集影响而泛化能力弱的问题(Breiman, 2001)。高维数据下随机森林产生的并行分类器在处理速度和分类精度都有较好的效果,但当数据噪声量过大时,随机森林仍然缺少去除多余噪声的能力,在分类过程中出现过拟合(Wang et al, 2018),特征个数过大,会产生过拟合问题,而特征个数不够,则会降低每个树的分类能力,从而增加了算法的错误率,针对该问题,该算法采用基于OOB(out of bag)误差的无偏估计进行随机选择最大特征数的确定。

3 结果与讨论

3.1 分割结果

使用ESP分割尺度工具时需要先确定分割的起始尺度,本文每隔10单位进行一次分割,目视对比分割效果和各分割尺度下的LV与ROC的曲

线变化,当分割尺度为 70 时,既可以较好地将盐碱地与其他地物分割开,又避免了内部的过分割现象。尺度分割 LV 与 ROC 变化如图 1 所示,可以看出当分割尺度为 71、97、123、132、152、166 时为峰值,分别使用这些尺度对影像进行分割,当分割尺度为 123 时,能较好地地区分盐碱地与周围地物类型。对于异质性相关因子权重的设置,采用单一参数设定法进行多次试验,发现增加近红外波段的权重,可以使盐碱地在该波段光谱信息更为丰富;将形状异质性与紧致度因子分别设定为 0.5 和 0.6,能够最大化体现目标的边界特征。

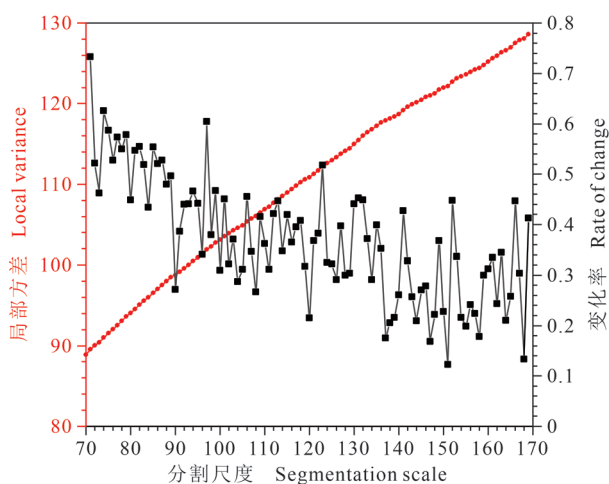


图 1 分割尺度效果图

Fig. 1 Effect of segmentation scale

3.2 特征优选

从目标对象纹理特征、光谱特征、形状特征以及遥感指数等自定义特征中选择适宜的特征数量和类型,可以提高分类的精度,减少数据的冗余计算。本文针对盐碱地的特征信息,构建了初始特征空间,在光谱特征与纹理特征中,每一个特征属性均在蓝、绿、红、近红外波段上进行特征构建,最大化保留每一个波段的特征信息,选择遥感指数 SAVI (土壤调节植被指数)、NDVI (归一化植被指数)、SI (盐度指数) 作为独立波段参与分类。该初始特征空间包含 93 个特征。初始特征空间中各属性数据量级不同,为防止数据数值之间差异过大而导致数据被吞噬问题,对各属性数据进行 Min-max 标准化处理,将各属性值限定至 $[-1, 1]$ 。

利用 Relief F 算法对初始特征空间进行降维,采取 Ranker 搜索策略对每一个特征进行权重计算并按顺序排列,得到的特征重要性排序如图 2 所

示。按排序可知前 9 个特征得分较高,依次为 HIS 变换、土壤调节植被指数、归一化植被指数、最大差分、近红外波段、红波段、盐度指数、蓝波段、绿波段;第 10—20 个特征处于中等得分,主要是灰度共生矩阵纹理特征以及部分形状特征,说明光谱特征与遥感指数在盐碱地提取中占有重要地位,纹理特征次之,形状特征最末。这是由于盐碱地的含盐量越高,在近红外区域反射率则越高,而裸露盐碱地多与盐生植被混合存在,因此纹理方向不定,纹理特征复杂。盐生植被的簇状生长,在土壤含盐量高的地方生长受抑制,土壤含盐量低则生长旺盛,盐碱地也受到其影响,形状呈簇状聚集,边缘形状弯曲多变。Relief F 算法虽然赋予了每个特征权重,但不能确定特征子集的数目,本文利用随机森林方法对数据集进行建模,通过得到不同特征数目下的分类总体精度 (overall accuracy) 与 Kappa 系数来确定最优特征数目。建模过程中对训练集采用十折交叉验证法进行训练,即将数据集分为十等份,将其中 9 份作为训练集,1 份作为验证集,直到每份数据都作为验证子集进行验证且验证 1 次。十折交叉验证法使得每个数据都参与了训练与测试两个环节,避免了模型的过度学习以及欠学习。由图 3 可知:随着特征变量数目逐渐增加,总体精度与 Kappa 系数也在快速增加,当特征变量达到 15 之后,曲线开始转变为波动状态,直到当特征数目为 40 时,精度达到峰值,分类总体精度达到 96%,Kappa 系数为 0.95,因此选择前 40 个特征作为优选特征,选择特征如表 2 所示。

针对 CFS 算法,采用全局最优算法 (best first) 作为搜索策略进行启发式搜索,进行特征预选,去除不相关变量。CFS 算法并不对每个特征变量进行排序,通过对特征子集的评估直接得到最优特征空间。通过 CFS 筛选后共有 17 个特征,结果如表 2 所示。通过比较可得两个算法所获得的特征中共有 14 个重合特征。基于 Relief F 得到的特征结果包括了 82% 的 CFS 筛选的特征,说明两种算法对重要特征均有较好的搜索效果,通过对比可以看出:各波段的光谱信息与自定义遥感指数在两种算法中都得到了保留,与上述重要性得分排序表现的结论相同,说明光谱信息在识别盐碱地类中的重要性,是区分其与其他地类的重要特征。CFS 算法在纹理特征与形状特征中相对于 Relief F 算法则约简了更多属性。

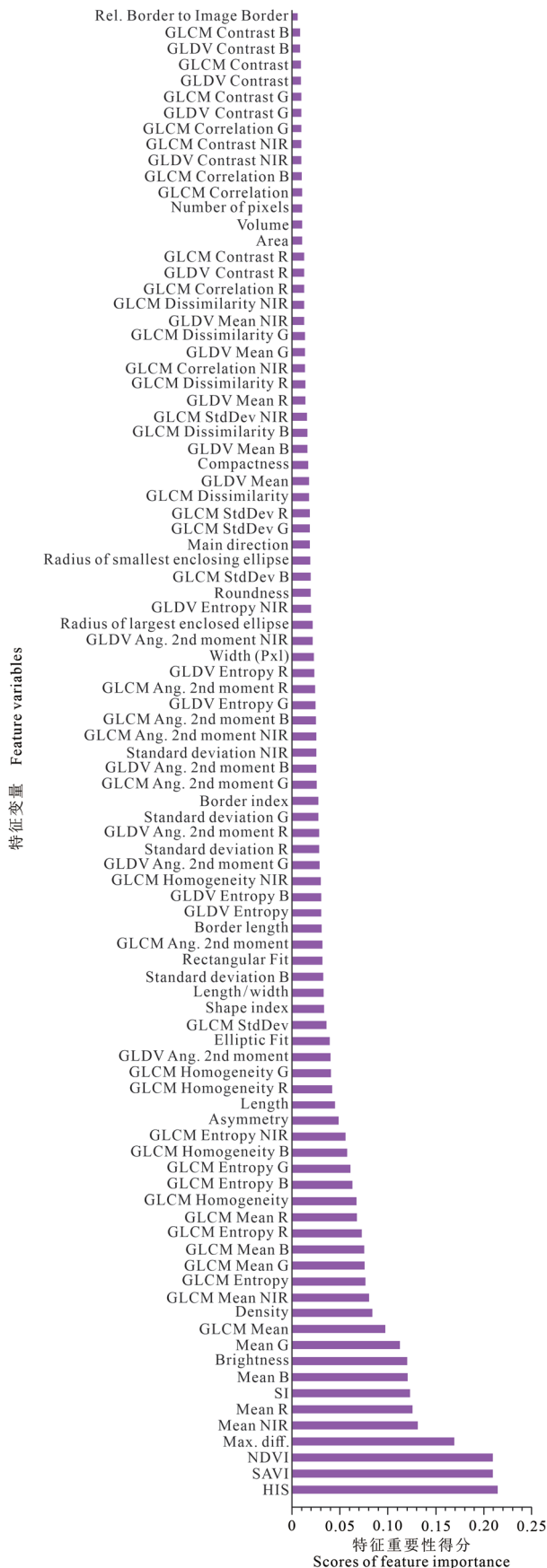


图2 特征重要性得分

Fig. 2 Scores of feature importance

3.3 分类结果及精度验证

通过 Relief F 与 CFS 算法筛选后的特征对研究区域进行随机森林分类, 通过选取的随机样本点利用解译标志以及 Google Earth 目视解译赋予样本数据属性, 利用生产者精度 (produce accuracy)、用户精度 (user accuracy) 统计不同算法下漏分误差与错分误差, 利用总体分类精度 (overall accuracy)、Kappa 系数评价总体分类效果。由表 3 可知: Relief F-RF 在盐碱地与盐化植被的分类上, 生产者精度与用户精度均有所提高, 但在其他类别上精度降低了 0.7% 与 1.6%, 原因是 Relief F 算法进行特征筛选时, 对于相关性较强的特征过滤效果不好, 导致冗余特征, 进而影响了分类性能; Relief F-RF 算法的总体精度提高了 1.1%。而 CFS 基于相关性的特征筛选算法则在分类效果有了明显提升, 相较于直接建立随机森林, 总体精度达到 83.7%, 提高了 7.4%, Kappa 系数为 0.74, 盐碱地的生产者精度提高了 8.9%, 用户精度提高了 6.8%。盐化植被的生产者精度提高了 10.4%, 用户精度提高了 12.2%。

上述结果表明: 基于 Relief F 与 CFS 对随机森林进行优化均能提高对盐渍化土地的提取精度, 未进行特征优选的随机森林算法提取精度最低, 表现出多维数问题对机器学习的分类性能存在一定的影响。Relief F 算法在特征中筛选出 40 个特征, 相较来说 CFS 算法只保留了 17 个特征, 却得到了更高的提取精度, 说明特征数目与精度不呈现正相关性, 特征之间的高相关性也会影响机器的学习能力 (李文杰等, 2020)。

不同算法下的分类结果以及局部细节如图 4 所示。通过对比遥感影像图, 可以看出 CFS-RF 算法对盐碱地识别精度更高。究其原因, 是受到土壤含盐量高低的影响, 盐渍化土地上会生长较为单一的植被如碱蓬、芦苇 (*Phragmites australis*) 等, 分布稀疏, 呈簇状聚集。植被的生长发育受到抑制, NDVI 值会比正常植被低, 在假彩色图像上呈现暗红色, 而盐渍化严重区域则会表现出亮斑现象。盐碱地与盐化植被在影像上通常呈交错分布, 盐渍化植被分布不均, 与盐碱地成为混合像元。对于盐碱地与盐化植被的混分现象, 通过多尺度分割以及 CFS 的特征筛选, 增大了类别之间的可分离性, CFS 优化后盐化植被的分类精度提升最多。有些盐碱地区域由于排水不畅、土壤湿度增大造成的地表反射率降低, 在影像上呈现暗色调, CFS-RF 对此也有很好的识别效果。

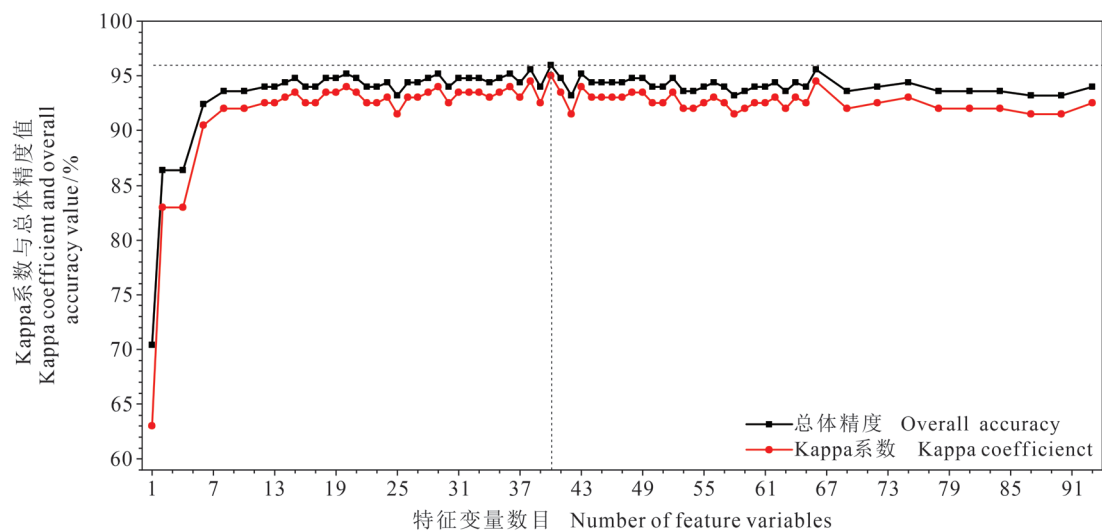


图 3 特征数目与精度关系

Fig. 3 Relationship between feature numbers and accuracy

表 2 Relief F 与 CFS 特征优选结果				
Tab. 2 Results of Relief F and CFS feature selection				
算法 Algorithm	光谱特征 Spectral features	形状特征 Shape features	纹理特征 Texture features	自定义特征 Custom features
Relief F	蓝、绿、红、近红 波段均值 Mean (B/G/R/NIR) 蓝波段标准差 Standard deviation B 最大差分 Max. diff. 亮度 Brightness	不对称性 Asymmetry 长度 Length 椭圆拟合性 Elliptic fit 矩形特征 Rectangular fit 边界长度 Border length 形状指数 Shape index 长宽比 Length/width 密度 Density	蓝、绿、红、近红波段、全方向灰度共生 矩阵中值 GLCM Mean (ALL/G/B/R/NIR)	HIS 变换 HIS 土壤调节植被指数 SAVI 归一化植被指数 NDVI 盐度指数 SI
			蓝、绿、红、近红波段、全方向灰度共生 矩阵熵 GLCM Entropy (ALL/G/B/R/NIR)	
			蓝、绿、红、近红波段、全方向灰度共生 矩阵协同性 GLCM Homogeneity (ALL/B/G/R/NIR)	
			全方向灰度共生矩阵标准差 GLCM StdDev ALL	
			全方向灰度共生矩阵角二阶矩 GLCM Ang. 2nd moment ALL	
			绿波段、全方向灰度共生矩阵归一化灰度 矢量角二阶矩 GLDV Ang. 2nd moment (ALL/G)	
			蓝波段、全方向灰度共生矩阵归一化灰度 矢量熵 GLDV Entropy (ALL/B)	
			全方向灰度共生矩阵协同性 GLCM Homogeneity ALL	
			全方向灰度共生矩阵角二阶矩 GLCM Ang. 2nd moment ALL	
			蓝波段灰度共生矩阵中值 GLCM Mean B	
			近红波段灰度共生矩阵标准差 GLCM StdDev NIR	
CFS	蓝、绿、红、近红 波段均值 Mean (B/G/R/NIR) 最大差分 Max. diff. 近红外标准差 Standard deviation NIR	宽度 Length 长宽比 Length/width 边界指数 Border index	全方向灰度共生矩阵协同性 GLCM Homogeneity ALL	HIS 变换
			全方向灰度共生矩阵角二阶矩 GLCM Ang. 2nd moment ALL	HIS
			蓝波段灰度共生矩阵中值 GLCM Mean B	土壤调节植被指数
			近红波段灰度共生矩阵标准差 GLCM StdDev NIR	SAVI
				归一化植被指数
				NDVI
				盐度指数

表 3 基于不同算法的分类精度统计
Tab. 3 Statistics of classification accuracy based on different algorithms

分类算法 Classification algorithms	类别 Class	生产者精度 Producer accuracy/%	用户精度 User accuracy/%	总体精度 Overall accuracy/%	Kappa 系数 Kappa coefficient
RF	盐化植被 Saline vegetation	66.7	48.5	76.3	0.63
	盐碱地 Saline-alkaline land	77.2	77.2		
	其他 Others	79.0	93.4		
Relief F-RF	盐化植被 Saline vegetation	70.8	54.0	77.4	0.64
	盐碱地 Saline-alkaline land	79.7	77.8		
	其他 Others	78.3	91.8		
CFS-RF	盐化植被 Saline vegetation	77.1	60.7	83.7	0.74
	盐碱地 Saline-alkaline land	86.1	84.0		
	其他 Others	84.6	97.6		

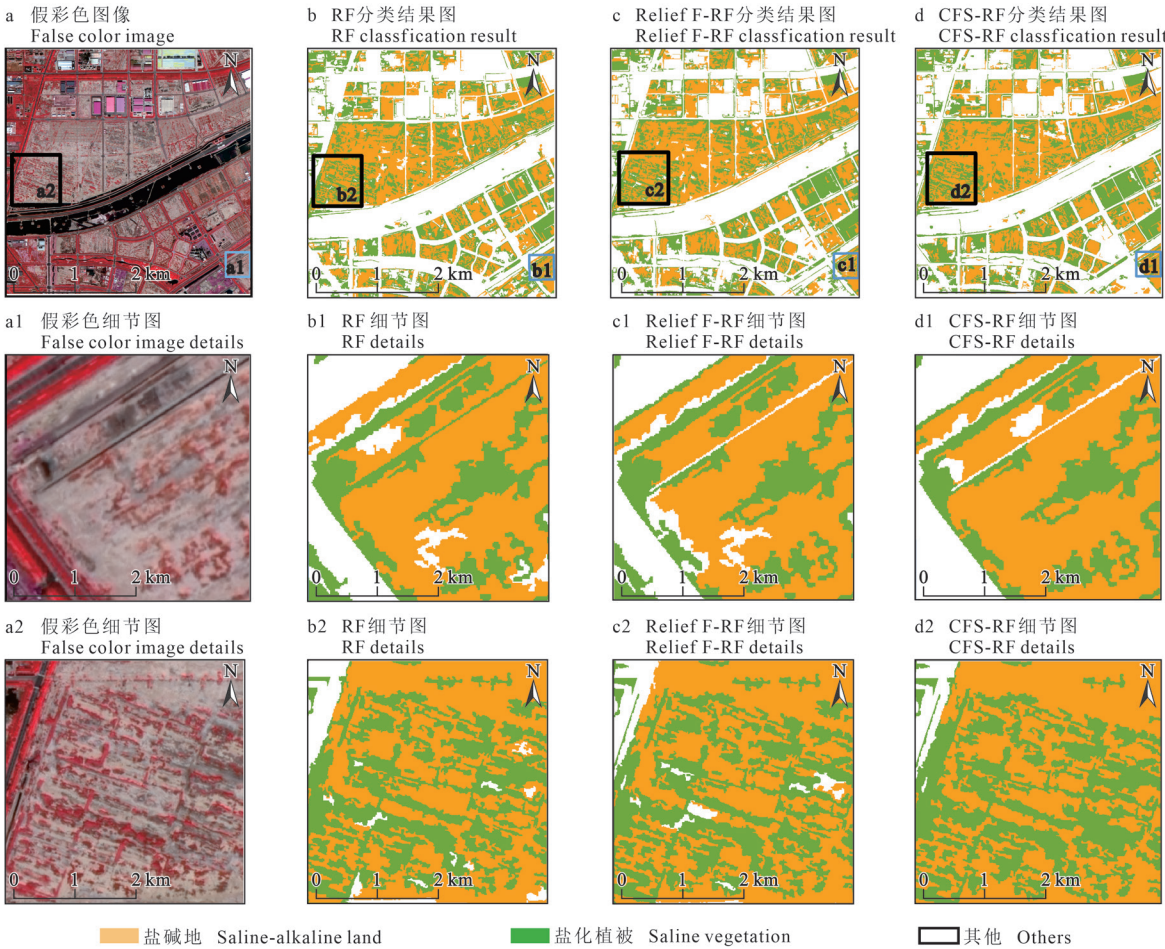


图 4 基于不同算法的分类结果比较
Fig. 4 Comparison of classification results based on different algorithms

4 结论

本文通过对 GF-6 遥感影像利用 FNEA 算法进行面向对象的多尺度分割, 确定适宜盐碱地提取的分割尺度, 利用 Relief F 与 CFS 算法进行特征筛选, 以此实现对随机森林的优化。得到结论如

下: (1) 确定了在 GF-6 高空间分辨率下盐碱地的分割尺度, 为盐渍化土地信息的提取提供参考依据; (2) 特征筛选可以过滤冗余数据, 提高机器学习的分类精度, CFS 算法对特征的过滤程度大于 Relief F 算法; (3) 本文提出的利用 CFS 算

法对随机森林进行优化,在盐碱地提取应用上有较好的分类效果,特征变量减至 17 个,且总体精度达到 83.7%,提高了 7.4%, Kappa 系数为 0.74;

(4) 对于高维属性数据特征复杂问题,本文提出的方法也可以适用于其他地类信息的提取,可以有效提高特征子集的数据质量与数据挖掘的效率。

参考文献

- 陈云浩,冯 通,史培军,等. 2006. 基于面向对象和规则的遥感影像分类研究 [J]. *武汉大学学报·信息科学版*, 31(4): 316–320. [Chen Y H, Feng T, Shi P J, et al. 2006. Classification of remote sensing image based on object oriented and class rules [J]. *Geomatics and Information Science of Wuhan University*, 31(4): 316–320.]
- 高仁强,陈亮雄,杨静学,等. 2020. 一种高分影像随机森林变化检测方法 [J]. *测绘科学*, 45(11): 130–138. [Gao R Q, Chen L X, Yang J X, et al. 2020. A method of random forest change detection based on high resolution image [J]. *Science of Surveying and Mapping*, 45(11): 130–138.]
- 何牧宇,周 晖. 2019. ReliefF-MFO 多标签特征选择算法 [J]. *计算机工程与设计*, 40(12): 3469–3473. [He M Y, Zhou H. 2019. ReliefF-MFO multi-label feature selection algorithm [J]. *Computer Engineering and Design*, 40(12): 3469–3473.]
- 姜 红,玉素甫江·如素力,热伊莱·卡得尔,等. 2017. 基于神经网络模型的干旱区绿洲土壤盐渍化评价分析 [J]. *地球信息科学学报*, 19(7): 983–993. [Jiang H, Yusufjiang Rusuli, Reyilai Kadeer, et al. 2017. Evaluation and analysis of soil salinization in the arid zones based on neural network model [J]. *Journal of Geo-Information Science*, 19(7): 983–993.]
- 李长龙,高志海,吴俊君,等. 2015. 基于分形网络进化分割和对象特征提取的 GF-1 卫星数据沙化土地分类识别研究 [J]. *干旱区资源与环境*, 29(11): 152–157. [Li C L, Gao Z H, Wu J J, et al. 2015. The sandy lands identification and classification of GF-1 based on FNEA and object features [J]. *Journal of Arid Land Resources and Environment*, 29(11): 152–157.]
- 李 晋,赵庚星,常春艳,等. 2014. 基于 HSI 高光谱和 TM 图像的土地盐渍化信息提取方法 [J]. *光谱学与光谱分析*, 34(2): 520–525. [Li J, Zhao G X, Chang C Y, et al. 2014. Land salinization information extraction method based on HSI hyperspectral and TM imagery [J]. *Spectroscopy and Spectral Analysis*, 34(2): 520–525.]
- 李文杰,郭晓雷,杨玲波,等. 2020. 基于 GF-6 卫星影像多特征优选的酿酒葡萄精准识别 [J]. *农业工程学报*, 36(18): 165–173. [Li W J, Guo X L, Yang L B, et al. 2020. Accurate recognition of wine grapes using multi-feature optimization based on GF-6 satellite images [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 36(18): 165–173.]
- 牛增懿,丁建丽,李艳华,等. 2016. 基于高分一号影像的土壤盐渍化信息提取方法 [J]. *干旱区地理*, 39(1): 171–181. [Niu Z Y, Ding J L, Li Y H, et al. 2016. Soil salinization information extraction method based on GF-1 image [J]. *Arid Land Geography*, 39(1): 171–181.]
- 孙宁青. 2010. 基于神经网络和 CFS 特征选择的网络入侵检测系统 [J]. *计算机工程与科学*, 32(6): 37–39, 117. [Sun N Q. 2010. A network intrusion detection system based on neural networks and the CFS-based feature selection [J]. *Computer Engineering & Science*, 32(6): 37–39, 117.]
- 王奕森,夏树涛. 2018. 集成学习之随机森林算法综述 [J]. *信息通信技术*, 12(1): 49–55. [Wang Y S, Xia S T. 2018. A survey of random forests algorithms [J]. *Information and Communications Technologies*, 12(1): 49–55.]
- 翁永玲,宫 鹏. 2006. 土壤盐渍化遥感应用研究进展 [J]. *地理科学*, 26(3): 369–375. [Weng Y L, Gong P. 2006. A review on remote sensing technique for salt-affected soils [J]. *Scientia Geographica Sinica*, 26(3): 369–375.]
- 徐存东,张 锐,王荣荣,等. 2018. 基于改进支持向量机的盐碱地信息精确提取方法研究 [J]. *灌溉排水学报*, 37(9): 62–68. [Xu C D, Zhang R, Wang R R, et al. 2018. An improved support vector machine method for estimating saline-alkali soil from remote sensing imagery [J]. *Journal of Irrigation and Drainage*, 37(9): 62–68.]
- 杨练兵,郑宏伟,罗格平,等. 2021. 基于遗传算法优化 BP 神经网络的土壤盐渍化反演 [J]. *地理与地理信息科学*, 37(2): 12–21, 37. [Yang L B, Zheng H W, Luo G P, et al. 2021. Retrieval of soil salinity content based on BP neural network optimized by genetic algorithm [J]. *Geography and Geo-Information Science*, 37(2): 12–21, 37.]
- 杨晓潇,王秀兰,王计平,等. 2019. 天津市滨海新区土壤水盐空间变异分析 [J]. *中国水土保持科学*, 17(3): 39–47. [Yang X X, Wang X L, Wang J P, et al. 2019. Spatial variation analysis of soil moisture and salinity in Tianjin Binhai New Area [J]. *Science of Soil and Water Conservation*, 17(3): 39–47.]

- 张 磊, 宫兆宁, 王启为, 等. 2019. 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*, 23(2): 313–326. [Zhang L, Gong Z N, Wang Q W, et al. 2019. Wetland mapping of Yellow River Delta wetlands based on multi-feature optimization of Sentinel-2 images [J]. *Journal of Remote Sensing*, 23(2): 313–326.]
- 张 萌. 2019. 面向对象框架下基于深度学习的高分辨率遥感影像分类 [D]. 昆明: 云南师范大学. [Zhang M. 2019. High-resolution remote sensing image classification based on object-oriented deep learning [D]. Kunming: Yunnan Normal University.]
- 张文博, 孔金玲, 杨园园, 等. 2021. 面向对象的旱区植被遥感精细分类研究 [J]. *测绘科学*, 46(1): 136–140, 183. [Zhang W B, Kong J L, Yang Y Y, et al. 2021. Fine vegetation classification of remote sensing in arid areas based on object-oriented method [J]. *Science of Surveying and Mapping*, 46(1): 136–140, 183.]
- Allbed A, Kumar L, Aldakheel Y Y. 2014. Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imageries: applications in a date palm dominated region [J]. *Geoderma*, 230/231: 1–8.
- Amaratunga D, Cabrera J, Lee Y S. 2008. Enriched random forests [J]. *Bioinformatics*, 24(18): 2010–2014.
- Benz U C, Hofmann P, Willhauck G, et al. 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3/4): 239–258.
- Breiman L. 2001. Random forests [J]. *Machine Learning*, 45(1): 5–32.
- Cui J T, Zhang X, Wang W S, et al. 2020. Integration of optical and SAR remote sensing images for crop-type mapping based on a novel object-oriented feature selection method [J]. *International Journal of Agricultural and Biological Engineering*, 13(1): 178–190.
- Drăguț L, Tiede D, Levick S R. 2010. ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data [J]. *International Journal of Geographical Information Science*, 24(6): 859–871.
- Elnaggar A A, Noller J S. 2009. Application of remote-sensing data and decision-tree analysis to mapping salt-affected soils over large areas [J]. *Remote Sensing*, 2(1): 151–165.
- Ivits E, Cherlet M, Tóth T, et al. 2013. Characterisation of productivity limitation of salt-affected lands in different climatic regions of Europe using remote sensing derived productivity indicators [J]. *Land Degradation & Development*, 24(5): 438–452.
- Jiang H, Rusuli Y, Amuti T, et al. 2019. Quantitative assessment of soil salinity using multi-source remote sensing data based on the support vector machine and artificial neural network [J]. *International Journal of Remote Sensing*, 40(1): 284–306.
- Kononenko I, Šimec E, Robnik-Šikonja M. 1997. Overcoming the myopia of inductive learning algorithms with RELIEFF [J]. *Applied Intelligence*, 7(1): 39–55.
- Li B Y, Wang Q W, Hu J L. 2011. Feature subset selection: a correlation-based SVM filter approach [J]. *IEEE Transactions on Electrical and Electronic Engineering*, 6(2): 173–179.
- Rodríguez J J, Kuncheva L I, Alonso C J. 2006. Rotation forest: a new classifier ensemble method [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10): 1619–1630.
- Sidike A, Zhao S H, Wen Y M. 2014. Estimating soil salinity in Pingluo County of China using QuickBird data and soil reflectance spectra [J]. *International Journal of Applied Earth Observation and Geoinformation*, 26: 156–175.
- Wang J Z, Ding J L, Yu D L, et al. 2019. Capability of Sentinel-2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China [J]. *Geoderma*, 353: 172–187.
- Wang Y S, Xia S T, Tang Q T, et al. 2018. A novel consistent random forest framework: bernoulli random forests [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8): 3510–3523.
- Ye Y M, Wu Q Y, Huang J Z, et al. 2013. Stratified sampling for feature subspace selection in random forests for high dimensional data [J]. *Pattern Recognition*, 46(3): 769–787.