

引用格式: Fang Panfei, Wang Leiguang, Xu Weiheng, *et al.* Decision Fusion Classification of Forest Dominant Tree Species in Shangri-La Area of Yunnan Province[J]. Remote Sensing Technology and Application, 2022, 37(3): 638-650. [方攀飞, 王雷光, 徐伟恒, 等. 云南香格里拉地区森林优势树种决策融合分类[J]. 遥感技术与应用, 2022, 37(3): 638-650.]

DOI: 10.11873/j.issn.1004-0323.2022.3.0638

云南香格里拉地区森林优势树种决策融合分类

方攀飞¹, 王雷光^{2,3}, 徐伟恒^{2,3}, 欧光龙¹, 代沁伶⁴, 李若楠¹

(1. 西南林业大学 林学院, 云南 昆明 650233;

2. 西南林业大学 大数据与人工智能研究院, 云南 昆明 650024;

3. 西南林业大学 森林生态大数据国家林业和草原局重点实验室, 云南 昆明 650024;

4. 西南林业大学 艺术与艺术学院, 云南 昆明 650233)

摘要: 基于 Google Earth Engine (GEE) 云计算平台, 协同 Sentinel-2 影像、WorldClim 生物气候数据、SRTM 地形数据、森林资源二类调查数据等数据, 以随机森林 (Random Forest, RF), 支持向量机 (Support Vector Machine, SVM) 和最大熵 (Maximum Entropy, MaxEnt) 3 种机器学习算法为组件分类器, 开展多源特征、多分类器决策融合的优势树种分类研究。通过 3 种组件分类器分别构建了两种串行集成和 3 种贝叶斯并行集成模型, 用于确定云南香格里拉地区 10 种主要优势树种的空间分布。分类结果显示: 3 个组件分类器的总体精度均低于 67.17%; 3 种并行集成方法总体精度相当, 约为 72%; 两种串行集成方法精度高于 78.48%, 其中 MaxEnt-SVM 串行集成方法获得最佳精度 (OA: 80.66%, Kappa: 0.78), 与组件分类器相比精度至少提高了 13.49%。研究表明: 决策融合方法在优势树种分类中比组件分类器精度更高, 并且有效改善了小样本树种的分类精度, 可用于大范围山区优势树种分类。

关键词: 优势树种; 机器学习; 决策融合; GEE

中图分类号: TP79 **文献标志码:** A **文章编号:** 1004-0323(2022)03-0638-13

1 引 言

优势树种分布信息是生物量评估、蓄积量评估、碳储量估算、栖息地质量评估的基础^[1-3], 也是森林管理部门有效管理和养护森林生态系统的关键^[4]。在传统森林资源调查中, 优势树种分布信息通过实地调查获得, 但受地形和天气等因素影响, 调查成本高。通过机器学习技术从遥感影像中获取树种空间分布信息成为一种高效便捷的方法^[2]。

随着遥感技术的发展, 利用遥感影像进行植被分类的研究层出不穷, 但是大区域山区树种遥感分

类仍然面临诸多困难。具体包括: ①多云多雨的气候条件导致难以获取高质量影像数据。②大范围大比例尺制图需要下载、存储和处理大量的数据, 本地单机系统处理效率低下。③复杂的生长环境和树种组成造成“异物同谱”和“同物异谱”现象, 仅使用光谱特征难以区分树种。④树种垂直地带性分布导致样本的空间分布和数量不均衡。不同树种分布面积相差较大, 样本数据数量与分布的不均衡会影响分类的准确性^[5-7]。

近 20 a 来, 集成学习方法被广泛应用于模式识

收稿日期: 2021-08-11; 修订日期: 2022-05-18

基金项目: 国家自然科学基金项目 (31860182, 32160369, 41571372, 31860181, 32060320), 云南省中青年学术和技术带头人后备人才项目 (2018HB026), 云南省基础研究计划面上项目 (202101AT070039)。

作者简介: 方攀飞 (1997—), 男, 云南宜良人, 硕士研究生, 主要从事资源与环境遥感研究。E-mail: 2862382468@qq.com

通讯作者: 王雷光 (1982—), 男, 安徽临泉人, 教授, 主要从事遥感图像分析与林业遥感应用相关研究。E-mail: wlgbain@126.com

别、图像处理的各个领域。在土地利用解译、医学和灾害预测和建模等多个方向的研究表明,集成学习可以更好地处理噪声数据以及不平衡数据,从而得到比基分类器更加精确的结果。Jafarzadeh等^[8]设计不同的分类场景来比较集成学习算法和基分类器对多光谱、高光谱和PolSAR 3类数据分类的性能,结果表明,对于不同类型的数据,集成学习算法分类性能比基分类器更强大。Rosa Aguilar等^[9]使用5个机器学习分类器绘制了马里南部小农耕作系统,基于加权投票策略的集成方法获得了较高分类精度,比最佳的基分类器精度高4.65%。Deepan等^[10]采用多层感知机、SVM和RF对PatternNet数据集中的8 000景影像分类,并采用多数投票方法进行集成,发现多数投票方法比单个分类器鲁棒性更高,分类结果更加准确。Kibria等^[7]通过集成多层神经网络和KNN算法的初始分类结果,设计了一种决策级融合模型,提高了心脏疾病的预测精度。Mahfuzur Rahman等^[5]用5种机器学习算法对孟加拉国全国范围内的不同类型洪水易发地区进行评估,结果表明集成方法预测洪水发生概率性能优于5种基础算法。总而言之,集成方法在提高分类准确性和可靠性方面展现出巨大的潜力^[11]。

在优势树种遥感识别中,运用多源数据融合和多分类器集成思路的研究尚不多见。鉴于此,本文基于GEE平台,探索结合机器学习和生态位模型、

高分辨率遥感数据和环境数据绘制大范围山区优势树种空间分布的分类方法。期望通过协同运用特征融合和决策融合,减少由于非最优特征、过拟合、训练样本不足等因素导致的误分类,以获得更稳健的分类结果。首先,采用特征融合与SVM、RF和MaxEnt 3个组件分类器获得初始分类结果;然后,设计决策融合规则,获得融合分类结果;最后,通过分类精度评价和分类结果的不确定分析,验证所提方法的有效性。

2 研究区与数据

2.1 研究区概况

香格里拉市隶属云南省迪庆藏族自治州,位于云南省西北部,其地理位置为 $99^{\circ}20' \sim 100^{\circ}19' E$ 、 $26^{\circ}52' \sim 28^{\circ}52' N$ 。如图1所示,研究区地势北高南低,最高海拔5 545 m,最低点海拔1 503 m,平均海拔3 459 m,立体气候明显,干湿季分明^[12]。地形和气候条件的不均匀性导致沿着适宜的环境条件形成各种植被群落。分布较广的优势树种有冷杉、云南松、高山松和栎类等。沿着环境梯度的空间分布模式以及均质斑块使得优势树种的分类制图成为可能。

2.2 数据

研究选择研究区内10种分布较广且具有代表性的优势树种进行分类。10种优势树种分别为:冷

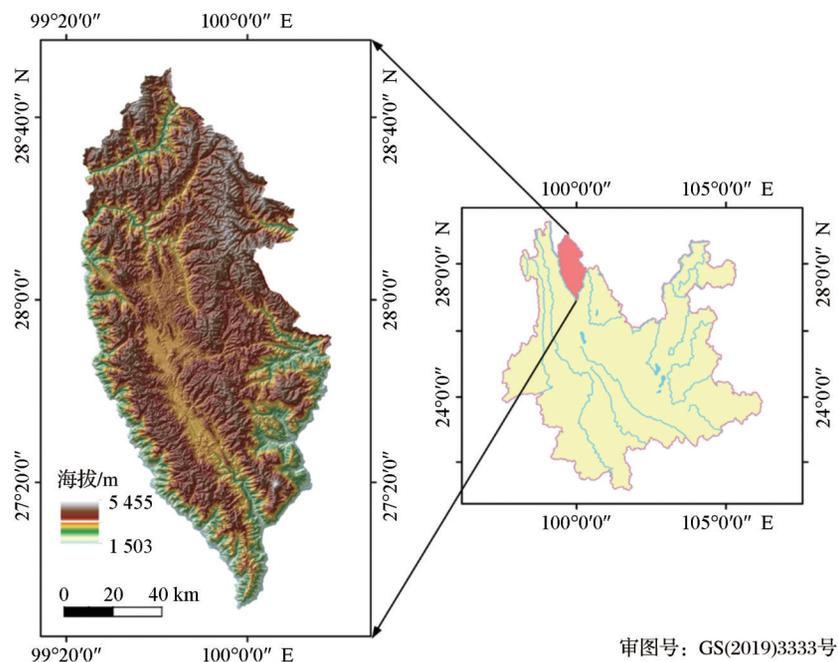


图1 研究区位置

Fig.1 Location map of the study area

杉(*Abies fabri*)、高山松(*Pinus densata* Mast.)、栎类(*Quercus acutissima*)、云南松(*Pinus yunnanensis*)、落叶松(*Larix gmelinii*)、桦类(*Betula L.*)、杨树(*Populus L.*)、柏木(*Cupressus funebris* Endl.)、华山松(*Pinus armandii* Franch.)和铁杉(*Tsuga chinensis* (Franch.) Pritz.)。

样本数据来源于森林资源二类调查数据,选取单一优势树种覆盖超过65%的纯林小班筛选样本数据。对选取的小班实施20 m向内缓冲以减少小班的边缘效应。然后用缓冲后的小班统计 Sentinel-2 波段 NIR 和 SWIR 波段反射率标准偏差,并剔除了标准差较大的不均匀小班。从筛选后的小班内生成 18 864 个样本点,其中训练样本 13 268 个,验证样本 5 596 个。由表 2 可以看出,高山松与铁杉的样本数量比例接近 9:1,样本数量的严重不平衡导致分类充满挑战。

表 2 优势树种样本数据

Table 2 Reference data for dominant tree species

优势树种	训练样本/个	测试样本/个
冷杉	1 937	834
高山松	3 125	1 320
栎类	1 686	696
云南松	1 474	662
落叶松	1 265	515
桦类	905	366
杨树	1 010	398
柏木	582	268
华山松	935	396
铁杉	349	141
总计	13 268	5 596

Sentinel-2 有 5 d 的高时间分辨率,包括从可见光(VIS)和近红外(NIR)波段到短波红外(SWIR)波段的 13 个波段,可见光的 R、G、B 和 NIR 波段的分辨率分别为 10 m,4 个红边波段分辨率为 20 m,轨道宽度为 290 km,这使得 Sentinel-2 适合大规模的森林类型和树种分类任务^[13-15]。

生物气候数据由(<http://www.worldclim.org>)网站上获得。如表 1 所示,此数据集包含 19 维空间分辨率为 1 km 的特征,分别代表年度趋势(年均温度、年降水量)、季节性(年度温差和降水量)以及极端或限制性环境因素(最冷和最热月份的温度,以及湿季和旱季的降水量)^[16-17]。

地形特征来自于 SRTM DEM(The Shuttle Radar Topography Mission Digital Elevation Model),

分辨率为 30 m,并从 SRTM DEM 中生成坡度(Slope)和坡向(Aspect)。

表 1 19 个生物气候变量

Table 1 Nineteen bioclimatic variables

变量	编码	分辨率/m
年均温	BIO1	1 000
昼夜温差月均值	BIO2	1 000
等温性	BIO3	1 000
气温季节性变化标准差	BIO4	1 000
最暖月最高温	BIO5	1 000
最冷月最低温	BIO6	1 000
气温年较差	BIO7	1 000
最湿季均温	BIO8	1 000
最干季均温	BIO9	1 000
最暖季均温	BIO10	1 000
最冷季均温	BIO11	1 000
年降水量	BIO12	1 000
最湿月降水量	BIO13	1 000
最干月降水量	BIO14	1 000
降水量的季节性变异系数	BIO15	1 000
最湿季降水量	BIO16	1 000
最干季降水量	BIO17	1 000
最暖季降水量	BIO18	1 000
最冷季降水量	BIO19	1 000

3 研究方法

本研究优势树种识别主要步骤如图 2 所示。首先,从数据集中提取特征;然后,筛选特征分别应用于组件分类器;在此基础上,对 3 个组件分类器的分类结果进行串行集成和贝叶斯并行集成两类决策融合,获得 10 m 分辨率的优势树种空间分布图;最后对结果进行评价分析。

3.1 特征融合

3.1.1 影像合成与特征提取

研究筛选 2018 年 12 月 1 日~2021 年 1 月 1 日,覆盖香格里拉地区、云量小于 30% 的 Sentinel-2 影像,共获得 382 景。将 382 景影像按采集时间划分为 4 个子时间序列。其中 3 月~5 月为春季,6 月~8 月为夏季,9 月~11 月为秋季,12 月至次年 2 月为冬季。其中,春季 71 景,夏季 51 景,秋季 86 景,冬季 174 景。受云雨影响,夏季可用影像数量最少。

将 4 个子时间序列影像以像素为单位按中值合成得到 4 景合成影像,基于 4 景合成影像分别提取 29 维光谱特征、50 维纹理特征,外加 3 维地形特征。如表 3 所示,合成的 4 景影像均有 83 维特征。光谱特征包括 3 类,第一类是合成影像的蓝色到短波红

外的 10 个波段 (Blue、Green、Red、RedEdge-1、Red-Edge-2、RedEdge-3、NIR、RedEdge-4、SWIR-1、SWIR-2); 第二类是植被指数, 包括三角植被指数 (Triangular Vegetation Index, TVI)^[18]、陆地叶绿素指数 (MERIS Terrestrial Chlorophyll Index, MTCI)^[19]、归一化火烧指数 (Normalized Burn Ratio, NBR)^[20]、反红边叶绿素指数 (Inverted Red-Edge Chlorophyll Index, IRECI)^[21]、综合叶绿素光谱指数 (Modified Chlorophyll Absorption Ratio Index, MCARI)^[22]、陆地表面水体指数 (Land Surface Water Index, LSWI)^[23]、归一化差异红边指数 (Normalized Difference Red Edge Index, NDRE)^[24]、叶绿素归一化指数 (Normalized Difference 750/705 Chl NDI,

Chl NDI)^[25]、红边位置指数 (Red-Edge Position Index, REP)^[26]、归一化植被指数 (Normalized Difference Vegetation Index, NDVI)^[27]、增强型植被指数 (Enhanced Vegetation Index, EVI)^[28]、修正简单植被指数 (Modified Simple Ratio Index, MSR)^[29]、红边比值植被指数 (RedEdge Ratio Index 2, RRI2)^[30], 以及基于合成的春秋两季影像构建的 TVI、NDRE、NDVI 的差值指数 (DIFTVI、DIFNDRE、DIFNDVI) 等。第三类是缨帽变换的绿色 (TCGRE) 和湿度 (TCWET) 分量。基于灰度共生矩阵 (GLCM) 对蓝色到短波红外 10 个波段分别计算 5 种纹理特征, 表 3 显示了提取的纹理特征及其计算公式, 其中 i 和 j 是矩阵中行和列数, $\mu_{i,j}$ 是均值, $\sigma_{i,j}$ 是标准差。

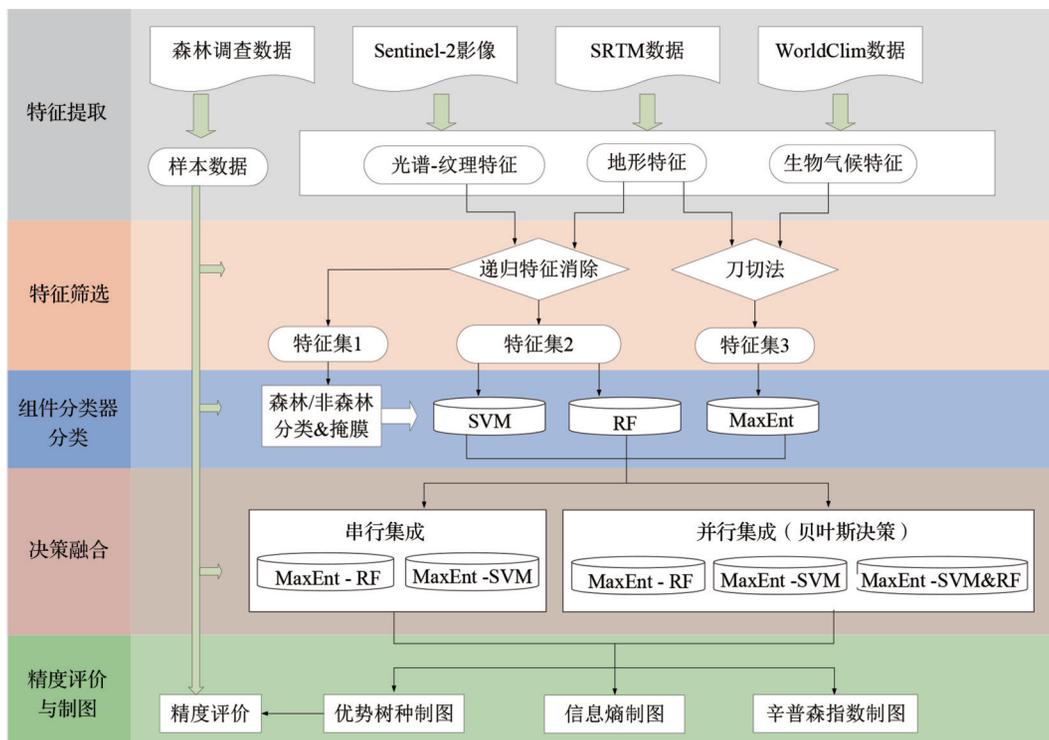


图 2 优势树种分类流程图

Fig.2 Workflow of dominant tree species classification

3.1.2 特征筛选

分离性较高的特征能提高分类性能, 而冗余和无用的特征会导致模型训练时间过长和过拟合等弊端, 造成“维数灾难”。因此, 筛选有效的特征是实现高精度分类的关键。

研究使用 19 维生物气候和 3 维地形特征作为 MaxEnt 模型输入, 首先, 利用软件内置的 Jackknife 折刀算法计算 22 维环境因子的贡献率, 再对 22 维环境因子进行 Spearman 相关性分析, 同属性因子中如相关性绝对值 ≥ 0.8 , 选择其中贡献率最高的因

子, 最终确定对优势树种分布预测贡献最大的环境因子。

对于 RF 和 SVM, 由光谱、纹理和地形特征构成的 83 维遥感因子作为输入, 首先对 83 维遥感因子进行归一化处理, 再用递归特征消除 (Recursive Feature Elimination, RFE) 算法^[31]选择最佳的森林与非森林分类以及优势树种分类特征。

3.2 决策融合

3.2.1 组件分类器

集成学习的核心是对多个基分类器初始分类

表3 用于优势树种分类的特征

Table 3 Features used for dominant tree species classification

特征	描述
Bands	Blue, Green, Red, Red_Edge1, Red_Edge2, Red_Edge3, NIR, Red_Edge4, SWIR_1, SWIR_2
TVI	$0.5 \times (120 \times (\rho_{RedEdge-1} - \rho_{Green}) - (200 \times (\rho_{Red} - \rho_{Green})))$
MTCI	$(\rho_{RedEdge-2} - \rho_{RedEdge-1}) / (\rho_{RedEdge-1} - \rho_{NIR})$
NBR	$(\rho_{NIR} - \rho_{SWIR-2}) / (\rho_{NIR} + \rho_{SWIR-2})$
IRECI	$(\rho_{RedEdge-3} - \rho_{Red}) / (\rho_{RedEdge-1} / \rho_{RedEdge-2})$
MCARI2	$((\rho_{RedEdge-2} - \rho_{RedEdge-1}) - 0.2 \times (\rho_{RedEdge-2} - \rho_{Green})) \times (\rho_{RedEdge-2} / \rho_{RedEdge-1})$
LSWI	$(\rho_{NIR} - \rho_{SWIR-1}) / (\rho_{NIR} + \rho_{SWIR-1})$
NDRE	$(\rho_{NIR} - \rho_{RedEdge-1}) / (\rho_{NIR} + \rho_{RedEdge-1})$
Chl NDI	$(\rho_{RedEdge-2} - \rho_{RedEdge-1}) / (\rho_{RedEdge-2} + \rho_{RedEdge-1})$
REP	$705 + 35 \times (0.5 \times (\rho_{Red} + \rho_{RedEdge-3}) - \rho_{RedEdge-1}) / (\rho_{RedEdge-2} - \rho_{RedEdge-1})$
NDVI	$(\rho_{NIR} - \rho_{RED}) / (\rho_{NIR} + \rho_{RED})$
EVI	$2.5 \times (\rho_{NIR} - \rho_{Red}) / (\rho_{NIR} + 6 \times \rho_{Red} - 7.5 \times \rho_{Blue} + 1)$
MSR	$(\rho_{NIR} / \rho_{RedEdge-1} - 1) / \sqrt{(\frac{\rho_{NIR}}{\rho_{RedEdge-1}}) + 1}$
RRI2	$\rho_{RedEdge-1} / \rho_{Red}$
DIFTVI	$TVI_{SPR} - TVI_{AUT}$ 春季的TVI减去秋季的TVI
DIFNDRE	$NDRE_{SPR} - NDRE_{AUT}$ 春季的NDRE减去秋季的NDRE
DIFNDVI	$NDVI_{SPR} - NDVI_{AUT}$ 春季的NDVI减去秋季的NDVI
TCGRE	Tasseled Cap Greenness
TCWET	Tasseled Cap Wetness
SAVG	$Sumaverage = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (i \times GLCM(i,j) + j \times GLCM(i,j))$
DIS	$Dissimilarity = \sum_{i=1}^m \sum_{j=1}^n i-j \times GLCM(i,j)$
VAR	$Variance = \sum_{i=1}^m \sum_{j=1}^n (j - \mu_j)^2 \times GLCM(i,j)$
CON	$Contrast = \sum_{i=1}^m \sum_{j=1}^n (i-j)^2 \times GLCM(i,j)$
CORR	$Correlation = \sum_{i=1}^m \sum_{j=1}^n \frac{\{i \times j\} GLCM(i,j) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y}$
Elevation	
Slope	
Aspect	

光谱特征

纹理特征

地形特征

结果实施决策融合来创建集成分类器,以获得更好的分类结果。一个成功的多分类器集成系统在很大程度上取决于组件分类器,高度多样性的组件分类器是构建有效的集成学习系统的关键,组件分类器越精确、差异越大,集成效果越好^[11]。本文选择SVM、RF和MaxEnt 3种差异较大且在土地利用、土地覆盖变化和森林场景分类中常获得较高精度的3个分类器进行集成。

SVM因计算速度快和泛化能力强等特点,被广泛用于植被分类。SVM通过非线性变换将输入的线性不可分空间转换成线性可分的高维特征空间,然后在这个高维特征空间中寻找最优超平面。最

优超平面不仅能够正确地对所有训练样本进行分类,而且能最大化最接近分类平面的点之间的距离,即能够以最大化分类间隔以分离不同的类别^[32]。Shao的研究结果表明,SVM在小的训练数据集上表现出比ANN和CART更好的性能^[33]。Matthew等用Sentinel-2数据和SVM对加利福尼亚州索诺马县的16种树种分类,分类精度为74.3%^[34]。

RF是一种bagging类型的集成算法,它利用多个决策树来合成预测。RF分类器通常比单个决策树具有更高的分类精度,当输入高维特征时,RF算法表现出较好的性能。Michael Schmidt通过陆地

卫星图像对作物/非作物分类时,比较了几种机器学习算法,结果证明,RF算法提供了更好的准确性^[35]。Agata Ho' scifo等使用Sentinel-2数据和RF算法对朝鲜高城郡的5种主要树种分类,精度达到80%^[36]。

MaxEnt是另一类被广泛用于确定树种空间分布的方法^[37]。它假设不完整的经验概率分布可以近似于一些环境因子约束的最大熵的概率分布,并且该分布近似于物种潜在地理分布^[38]。即使样本的尺寸和空间间隔小,MaxEnt也能得到稳定可靠的预测精度。多项研究表明MaxEnt模型的预测结果接近真实分布^[39-40]。

3.2.2 决策融合方法

设计合适的决策融合规则可以提高集成模型的性能,本文采用并行集成和串行集成两种范式集成3个组件器,如图3所示。

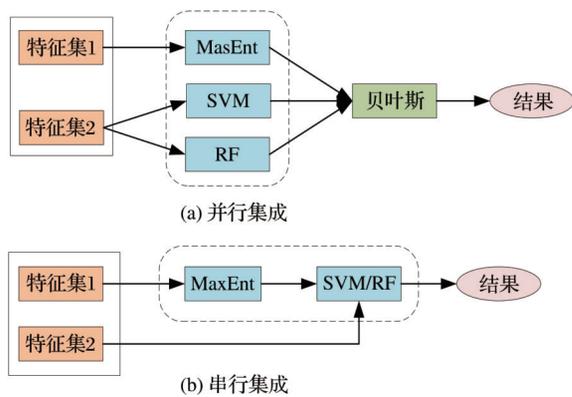


图3 决策融合方法示意图

Fig.3 Schematic diagram of decision fusion method

第一种范式是并行模式。使用3个组件分类器进行单独分类,然后基于贝叶斯决策思想进行决策融合。在极大后验分析的框架下,分类问题可以描述为:已知观测特征影像Z,类别集合Ω,像素位置集合S,寻找一个全局类别标记Y*,使得后验概率P(Y|Z)取得最大值,即

$$Y^* = \underset{Y}{\operatorname{argmax}} (P(Y|Z)) \quad (1)$$

根据贝叶斯准则,求解Y*等价于最小化后两部分能量之和,即

$$Y^* = \underset{Y}{\operatorname{argmin}} (E_f + E_l) \quad (2)$$

其中: $E_l = \sum_{i \in S} \log(P(y_i))$ 和 $E_f = \sum_{i \in S} \log(P(z_i|y_i))$, E_l 和 E_f 分别代表先验能量和类条件能量。在此框架下,令 $P(y_i) = I_{i,y_i}^{ENM}$,即将生态位模型预测出的像素

i 属于给定类别 $y_i \in \Omega$ 的概率 I_{i,y_i}^{ENM} 作为先验概率;令 $\log(P(z_i|y_i)) = \sum_j w_j I_{i,y_i}^{ML_j} I_{i,y_i}^{ML_j}$,且 $\sum_j w_j = 1$,即分类器获得的类别概率 $I_{i,y_i}^{ML_j}$ 线性加权,实现不同机器学习分类方法的融合,作为类条件概率,并有最终融合目标函数:

$$Y^* = \underset{Y}{\operatorname{argmax}} \left(\sum_{i \in S} \log(I_{i,y_i}^{ME}) + \sum_{i \in S} \log(I_{i,y_i}^{ML_j}) \right) \quad (3)$$

实现多分类器结果与生态位模型预测树种分布信息的决策融合。以MaxEnt的分类结果作为先验概率,SVM、RF的分类结果以及SVM和RF的分类结果的加权求和作为后验概率,可以构建MaxEnt-SVM、MaxEnt-RF和MaxEnt-SVM&RF三类并行模型。

第二种集成范式是串行模式。以顺序模式执行的分类器链的算法,即上一个分类器的输出是链中下一个分类器的输入,该方法将MaxEnt的分类结果作为RF和SVM的输入,MaxEnt算法使用了多种生态环境因子,RF和SVM使用了光谱和纹理特征,实现特征级与决策级融合,受益于分类器链中所涉及的每种算法,得到最终分类结果。以MaxEnt的分类结果分别作为SVM和RF的输入,可构建MaxEnt-SVM和MaxEnt-RF两类串行模型。

3.3 分类结果评价方法

3.3.1 精度评价

为了评价组件分类器和集成方法的分类性能,使用总体精度(Overall Accuracy, OA)、用户精度(User's Accuracy, UA)、生产者精度(Producer's Accuracy, PA)、卡帕系数(Kappa)和F-Score(F1)5个指标描述分类准确率和类别间的混淆程度。

3.3.2 分类空间不确定性度量

由于分类精度的度量均基于有限的测试样本,难以反映整个研究区的分类精度的情况,因此引入辛普森指数和信息熵进一步分析分类结果的不确定性。

辛普森指数(Simpson Diversity Index, SI)^[41]的定义如下:

$$SI = 1 - \sum_{i=1}^{i=k} p_i^2 \quad (4)$$

其中: p_i 是某像素位置的分类预测为*i*类的分类器数量与总分类器数量的比例, k 是类别数量。SI用来评价多个分类结果的差异性,指数高代表分类的一致性低,不确定性大;指数低分类器之间的一致性高,分类结果可信度较高。本文使用5个集成模型计算SI。可以得到每一个像素的SI值从0(当所有分类器都把该像素分给*i*类)到0.8(所有分类器把

这个像素分成了不同的类)。

分类概率的信息熵 H 描述为

$$H(I) = -\sum_{i=1}^{i=n} P_i \log_2(P_i) \quad (5)$$

其中: p_i 是给定像素位置预测为 i 别的概率, n 是类别的数量。像素分类的信息熵 H 越大,则分类的不确定性越大。本文用精度最佳的模型来计算信息熵,当一个像素被划分为给所有类别的概率一样时熵值最大;而划分为某一类别概率接近1,划分其他类别的概率趋近0时熵值最小。

4 结果与讨论

4.1 特征重要性评价

结合 Jackknife 折刀算法和相关性分析对环境进行筛选得到对树种分布有较大影响的变量,利用筛选得到的主导环境因子对各优势树种进行分布

预测,预测结果与使用全部22维环境因子预测结果相比AUC值变化很小,说明筛选后的主导因子对于物种的解释力较强。表4展示了对于10种优势树种分布影响较大的特征变量及其贡献率,贡献率越大表示优势树种分布影响越大。由表4可以看出,Elevation对于多个树种都是贡献率最高的环境因子,特别是分布在高海拔区域的冷杉和高山松,贡献率均超过92%,对于落叶松的贡献率也达到了68.4%。极高的海拔限制有利于将3类树种与其他树种区分开。对于柏木和铁杉两类树种,最湿月降水量是贡献最高的因子,对于云南松贡献率最高的因子则是最冷季均温。对于各树种的主导因子有Slope,并且Aspect、BIO02和BIO03是多个树种的主导因子,表明这些树种的空间分布都受到地形因素和气温的影响。

表4 各优势树种分布的主导环境影响因子及贡献率

Table 4 Contribution rate of environmental factors affected on the distribution of each dominant tree species

树种	因子/贡献率(%)
冷杉	Elevation(93), BIO07(4), Slope(1.2), BIO02(1.1), BIO15(0.5), BIO19(0.2),
高山松	Elevation(92.3), BIO04(3.5), BIO02(2.6), Slope(0.8), BIO03(0.5), BIO15(0.2), Aspect(0.1)
栎类	Elevation(58.2), Slope(19), Aspect(9), BIO02(7.6), BIO12(4.8), BIO14(1.4)
云南松	BIO11(79.9), BIO13(8.5), BIO03(3.7), Slope(2), BIO12(1.8), BIO14(1.4), Aspect(1.4), BIO14(1.2)
落叶松	Elevation(68.4), BIO04(13.3), BIO02(11.6), BIO15(3.2), Aspect(3.1), BIO03(0.2), Slope(0.1)
桦类	Elevation(54.9), Slope(19.7), BIO15(15.6), BIO04(3.1), BIO02(3), Aspect(2.8), BIO14(1),
杨树	Elevation(50.2), BIO13(27.9), Slope(13.7), BIO03(2.2), BIO17(2.1), BIO15(2.1),
柏木	BIO13(58.6), Elevation(21.5), BIO01(5.7), BIO03(5), BIO19(3.4), Slope(3.1), Aspect(2.6)
华山松	Elevation(40.2), BIO05(27.3), BIO12(13), BIO15(7.2), Slope(5.9), Aspect(3.6), BIO17(1.7), BIO03(1.1)
铁杉	BIO13(61.3), BIO11(21.7), BIO07(8.3), BIO014(3.6), BIO02(1.9), BIO19(1.3), Aspect(1.2), Slope(0.5), BIO03(0.2)

表5 不同季节合成影像分类结果

Table 5 Classification results in terms of OA and Kappa for different seasonal image composites

季节	OA/%	Kappa
春季	68.28	0.63
夏季	69.55	0.65
秋季	67.96	0.62
冬季	66.76	0.62

对于遥感因子,首先,选择SVM分类器对春、夏、秋、冬4景影像的83个遥感因子分别进行分类,确定最佳的分类季相,表4展示了分类结果。当使用夏季影像时获得的分类精度最高,OA为69.55%,Kappa为0.65;当使用冬季影像时取得的精度最低,OA精度为66.76%,Kappa为0.51。虽然过去云后夏季影像最少,却是区分树种最佳的季相,这与 Salvatore Praticò 等的研究结论一致^[42]。然

后,评价夏季 83 维特征的重要性得分,如图 5 所示,高程是绘制优势树种空间分布最重要的变量,冷杉、高山松和云南松分布在不同海拔梯度,分类结果表明三者之间混淆程度较低。许多研究表明加入海拔、坡向等辅助数据可以提高植被分类的精

度^[15,43]。当然,其重要程度是根据研究区而定,一些研究的研究区地形平坦,多个树种分布在相同的地形条件,会使得地形特征在模型中解释力变小。光谱特征中,春季和秋季两个季节的差值指数的重要性得分较高,说明在香格里拉地区,各树种生长和衰

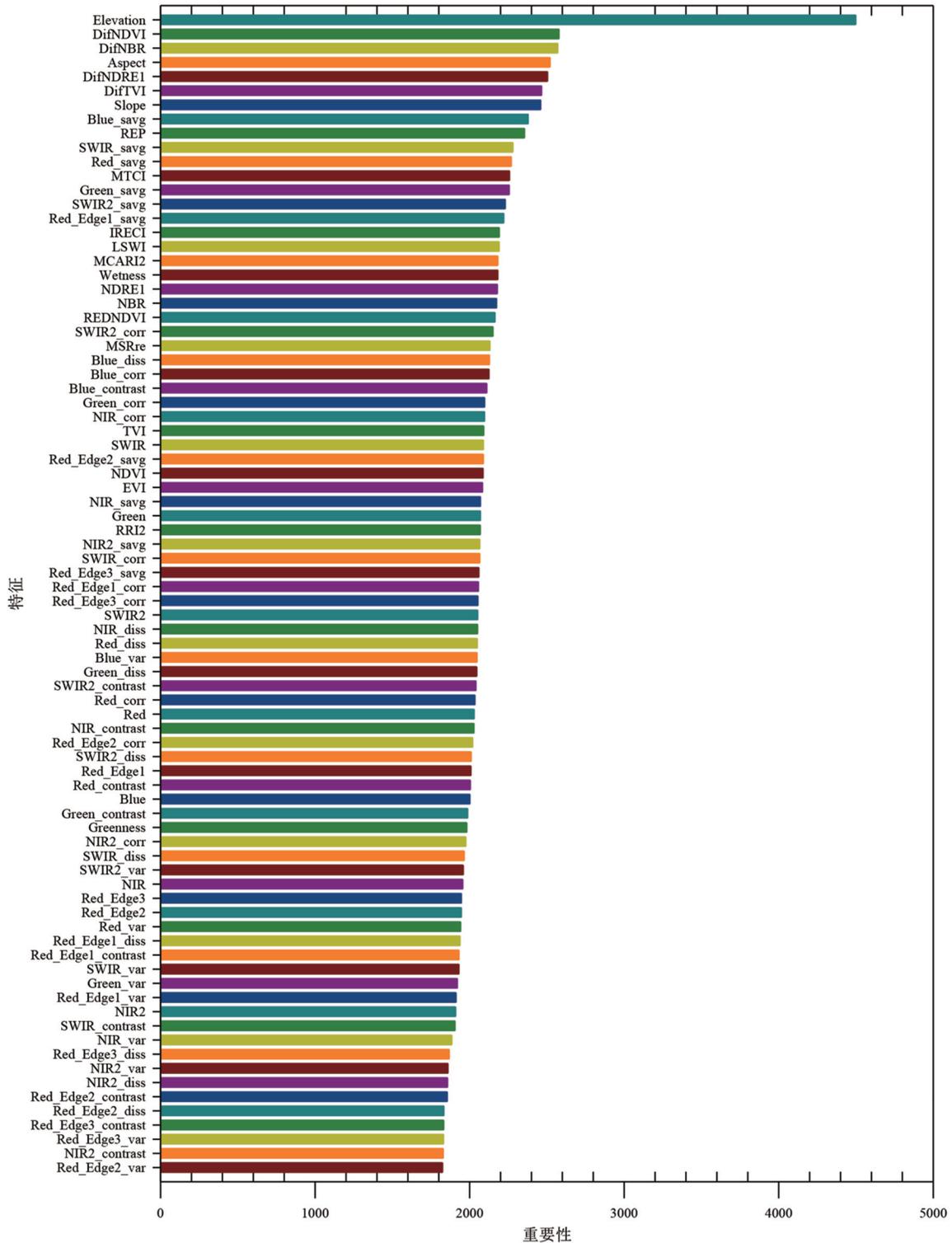


图 5 特征的重要性排序

Fig.5 Importance sorting of features

老模式沿海拔梯度的变化而表现出光谱异质性,这有助于区分各种树种。其次是基于红边波段构建的植被指数,如REP、MTCI、IRECI、LSWI、MCARI2等,说明了红边波段在树种制图中的重要性。纹理特征中由各波段计算的SAVG重要性得分值较高。

得到各维特征的重要性得分,本文对比了不同特征组合的对分类精度的影响,结果如图6所示。仅用光谱特征分类精度最低,SVM和RF的OA分别为54.89%和52.50%。加入地形特征对提升分类精度贡献最大,总体精度提高约10%~14%,表明在地形复杂的山区,树种的分布受到地形的影响较大。纹理特征的加入,精度提高1%~7%,证明了纹理特征的重要性。MaxEnt使用环境因子的精度高于使用光谱特征和纹理特征的SVM和RF,表明虽然生物气候特征空间分辨率差,但仍有利于树种分类。此外,用83维特征的分类精度(OA:69.55%)比采用RFE降维后50维特征的分类精度(OA:68.73%)略高,但RFE降维后的特征集只有50维,减少了计算负担。所以本文选择RFE降维后特征组合作为SVM和RF的输入特征。

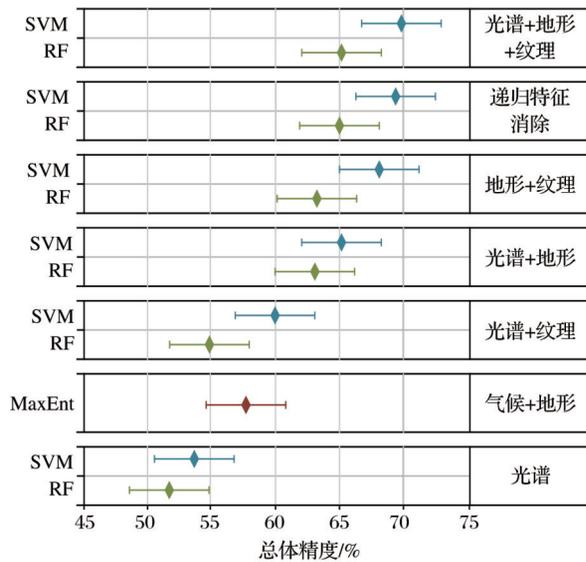


图6 不同分类器和特征组合分类结果

Fig.6 Classification results in terms of overall accuracy for different classifiers and feature combinations

4.2 决策融合方法在优势树种分类中的对比

研究对比了3个组件学习器与5类决策融合模型的分类精度。图7显示了各种模型的OA和Kappa。从图中可以看出,决策融合模型分类精度均优于组件分类器,证明了串行集成和并行集成两种范式都可以提高分类精度。与之前的研究一致,多样

性和准确性是构建集成模型的两个关键^[44-45],组件分类器间良好的性能和相互间的差异使得决策融合模型分类精度较之组件学习器精度提高了5%~21%。

决策融合模型中串行集成比并行集成的精度更高,最佳分类精度(OA:80.66%,Kappa:0.78)由MaxEnt-RF串行集成分类器得到,这与现有的在小区域使用高光谱或多光谱图像进行分类的研究相当,甚至高于部分研究的精度^[10,15,36,46-49]。这一精度表明,在区域尺度上,考虑植被生长的环境信息和遥感数据,通过集成机器学习和生态位模型可以绘制详细的优势树种空间分布图,且方法普适性较高,可以应用于林业清查、环境监测和碳循环估计等应用。

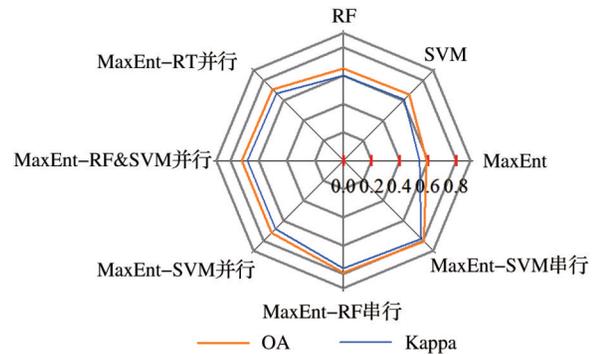


图7 不同分类方法精度对比

Fig.7 Comparison of accuracy of different classification methods

值得一提的是,决策融合方法显著改善了小样本类别的分类精度。图8显示每个类别的UA、PA和F1,可以看出每一个类别最佳的UA、PA和F1都是由决策融合方法提供的。特别对于小样本类别如杨树、华山松、柏木和铁杉四类,决策融合方法得到的精度明显高于组件学习器。大量研究也表明集成方法可以缓解数据集中类不平衡问题^[6,50-51],本文的分类结果也证实这一说法,对于大样本类别SVM和RF提供了更高的F1,而对于小样本类别MaxEnt提供了比SVM和RF更高的F1,当三者进行决策融合时能同时提高大样本类别和小样本类别的精度。

4.3 分类不确定性分析

辛普森指数图和信息熵图是对分类不确定性的可视化表达。集成方法产生了类分配不确定性信息,辛普森指数图评估了5种决策融合模型在每个像素上的分类一致性,见图9(a),而信息熵图是

UA	0.69	0.67	0.33	0.84	0.58	0.44	0.33	0.89	0.78	0.79	MaxEnt
	0.73	0.56	0.61	0.76	0.62	0.61	0.51	0.75	0.65	0.94	RF
	0.77	0.57	0.57	0.80	0.62	0.65	0.45	0.84	0.61	0.59	SVM
	0.76	0.70	0.45	0.80	0.69	0.72	0.47	0.92	0.86	0.83	MaxEnt-RF并行
	0.78	0.66	0.56	0.83	0.69	0.75	0.55	0.89	0.84	0.88	MaxEnt-SVM并行
	0.78	0.64	0.59	0.82	0.69	0.77	0.57	0.87	0.85	0.90	MaxEnt-RF&SVM并行
	0.80	0.73	0.71	0.81	0.73	0.78	0.76	0.91	0.93	0.87	MaxEnt-RF串行
	0.82	0.76	0.66	0.82	0.73	0.80	0.78	0.94	0.92	0.87	MaxEnt-SVM串行
	PA	0.63	0.39	0.52	0.63	0.64	0.63	0.62	0.71	0.77	0.67
0.83		0.80	0.42	0.82	0.68	0.46	0.25	0.79	0.31	0.21	RF
0.81		0.81	0.53	0.82	0.70	0.55	0.30	0.78	0.18	0.18	SVM
0.81		0.70	0.52	0.82	0.71	0.57	0.60	0.82	0.66	0.67	MaxEnt-RF并行
0.81		0.78	0.61	0.85	0.74	0.59	0.55	0.83	0.43	0.62	MaxEnt-SVM并行
0.84		0.79	0.58	0.84	0.75	0.58	0.50	0.83	0.40	0.57	MaxEnt-RF&SVM并行
0.86		0.83	0.48	0.88	0.78	0.61	0.66	0.96	0.85	0.95	MaxEnt-RF串行
0.83		0.81	0.54	0.87	0.80	0.69	0.77	0.96	0.89	0.94	MaxEnt-SVM串行
F1		0.66	0.49	0.40	0.72	0.61	0.52	0.43	0.79	0.77	0.73
	0.78	0.66	0.50	0.79	0.65	0.52	0.34	0.77	0.42	0.34	RF
	0.79	0.67	0.54	0.81	0.66	0.60	0.36	0.81	0.28	0.28	SVM
	0.78	0.70	0.48	0.81	0.70	0.64	0.53	0.87	0.75	0.74	MaxEnt-RF并行
	0.79	0.72	0.58	0.84	0.71	0.66	0.55	0.86	0.57	0.73	MaxEnt-SVM并行
	0.81	0.71	0.58	0.83	0.72	0.66	0.53	0.85	0.54	0.70	MaxEnt-RF&SVM并行
	0.83	0.78	0.57	0.84	0.75	0.68	0.71	0.93	0.89	0.91	MaxEnt-RF串行
	0.82	0.78	0.59	0.84	0.76	0.74	0.77	0.95	0.90	0.90	MaxEnt-SVM串行
		冷杉	高山松	栎类	云南松	落叶松	桦类	杨树	柏木	华山松	铁杉

图 8 不同分类器分类精度

Fig.8 Classification accuracy of different classifiers

基于精度最高的模型计算得到,见图 9(b),二者在空间有着很好的一致性。事实上,这二者信息可以体现分类结果的可信度,是对标准精度评估的补充,可用于分类后处理和分析^[52]。

从图 9(c)中可以看出,辛普森指数图和信息熵图显示分类精度高的区域在海拔较高的同质林分高、森林和冠层覆盖率高和优势树种丰度聚集分布地区,这些区域有较大的原生森林斑块,这些区域的特点是特征相对纯净。这些结论与 Ewa Grabska 等人一致^[43]。分类精度差的区域位于海拔过渡区,这些环境限制较低,多个树种的生境存在重叠,树种组成异质程度较高。

类别间的混淆程度表明:生态幅较窄,生境限制较大的树种与生态幅宽,对环境忍耐度大的树种相比分类精度更高。如冷杉、高山松和落叶松分布在高海拔区域,多个分类器得到的 UA、PA 和 F1 均较高,而栎类和杨树生态幅较广和多个树种有着相似生态驱动因素,容易和其他类别产生混淆,导致分类精度较低。

4.4 优势树种制图

MaxEnt-SVM 串行集成分类结果如图 9(d)所示。其中,多数优势树种都有自己的集中区域,表明香格里拉地区森林立地条件具有明显异质性,如冷杉和柏木分布在海拔较高的地区,而高山松和云南松类占据海拔较低的地区,云南松在南部的森林边缘地区呈现“V”形分布,落叶松集中在香格里拉的中部,桦类分布在中南部的中高海拔区域,栎类分布则较为离散。10 个优势树种中,冷杉分布面积最大,其次是高山松,分别占森林总面积的 25% 和 23%,栎类与云南松面积相近,分别占森林总面积的 13% 和 12%,落叶松与柏木面积也相近,分别占森林总面积的 9% 和 8%。

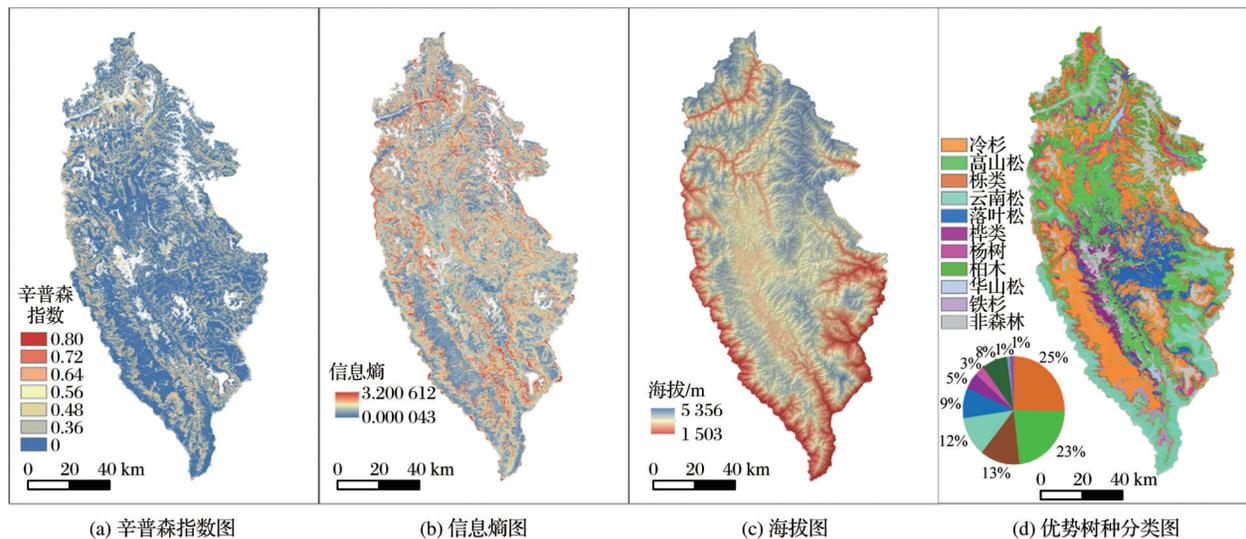


图 9 分类结果及不确定性图

Fig.9 Classification results and uncertainty diagrams

5 结 论

为解决山区优势树种分类面临的挑战和难题,本文使用机器学习模型和生态位模型建模、高分辨率 Sentinel-2 数据和降雨、气温和地形等环境因子的组合来绘制香格里拉地区 10 种优势树种空间分布图。该方法以 RF、SVM 和 MaxEnt 作为组件分类器,然后通过决策融合构成集成分类器,实现特征级和决策级融合。

主要结论有:①决策融合方法比组件分类器具有更高分类精度并明显改善了小样本类别的分类精度,其中基于 MaxEnt-RF 串行集成的分类器性能最优,OA 为 80.66%,Kappa 为 0.78,证明决策融合方法在大范围山地地区优势树种分类中的有效性。②夏季是研究区优势树种分类最有价值的物候期,地形特征对于山区的树种分类的有突出作用。③使用多源数据有助于提高优势树种分类精度,研究区垂直地带性气候变化明显,虽然生物气候变量分辨率为 1 km,但对于表征树种分布仍起到重要作用。在高海拔地区,由于生境限制林分同质性较高,还有优势树种丰度聚集的区域,优势树种识别效果较好,而容易错分的区域位于海拔过渡区。

参考文献 (References):

- [1] Waser L T, Ginzler C, Kuechler M, *et al.* Semi-automatic classification of tree species in different forest ecosystems by spectral and geometric variables derived from Airborne Digital Sensor (ADS40) and RC30 data[J]. *Remote Sensing of Environment*, 2011, 115(1): 76-85.
- [2] Ke Y, Quackenbush L J, Im J. Synergistic use of QuickBird multispectral imagery and LiDAR data for object-based forest species classification[J]. *Remote Sensing of Environment*, 2010, 114(6): 1141-1154. DOI:1110.1016/j.rse.2010.1101.1002.
- [3] Puletti N, Chianucci F, Castaldi C. Use of Sentinel-2 for forest classification in mediterranean environments[J]. *Annals of Silvicultural Research*, 2018, 42: 32-38.
- [4] Dalponte M, Bruzzone L, Gianelle D. Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data[J]. *Remote Sensing of Environment*, 2012, 123: 258-270. DOI:210.1016/j.rse.2012.1003.1013.
- [5] Rahman M, Chen N, Elbeltagi A, *et al.* Application of stacking hybrid machine learning algorithms in delineating multiple flooding in Bangladesh[J]. *Journal of Environmental Management*, 2021, 295: 113086.
- [6] Yang K, Yu Z, Wen X, *et al.* Hybrid classifier ensemble for imbalanced data [J]. *IEEE transactions on neural networks and learning systems*, 2019, 31(4): 1387-1400.
- [7] Kibria H B, Matin A. An efficient machine learning-based decision-level fusion model to predict Cardiovascular disease [C]// *Proceedings of the International Conference on Intelligent Computing & Optimization*, F, 2020. Springer.
- [8] Jafarzadeh H, Mahdianpari M, Gill E, *et al.* Bagging and boosting ensemble classifiers for classification of multispectral, hyperspectral and PolSAR data: A comparative evaluation [J]. *Remote Sensing*, 2021, 13 (21) : 4405. DOI: 4410.3390/rs13214405.
- [9] Aguilar R, Zurita-Milla R, Izquierdo-Verdiguier E, *et al.* A cloud-based multi-temporal ensemble classifier to map smallholder farming systems [J]. *Remote Sensing*, 2018, 10(5) : 729. DOI:710.3390/rs10050729.
- [10] Deepan P, Sudha L. Scene classification of remotely sensed images using ensembled machine learning models [M]. *Machine Learning, Deep Learning and Computational Intelligence for Wireless Communication*. Springer. 2021: 535-550.
- [11] Du P, Xia J, Zhang W, *et al.* Multiple classifier system for remote sensing image classification: A review [J]. *Sensors*, 2012, 12(4): 4764-4792. DOI:4710.3390/s120404764.
- [12] Farong S. Current status and characteristics of forest resources in Shangri-la County [J]. *Journal of West China Forestry Science*, 2008, 37(1): 124-128. [宋发荣. 香格里拉县的森林资源及其特点分析 [J]. *西部林业科学*, 2008, 37(1): 124-128.]
- [13] Wessel M, Brandmeier M, Tiede D. Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on Sentinel-2 data [J]. *Remote Sensing*, 2018, 10(9): 1419. DOI:1410.3390/rs 10091419.
- [14] Grabska E, Hostert P, Pflugmacher D, *et al.* Forest stand species mapping using the Sentinel-2 time series [J]. *Remote Sensing*, 2019, 11(10): 1197. DOI:1110.3390/rs11101197.
- [15] Hościło A, Lewandowska A. Mapping forest type and Tree Species on a regional scale using multi-temporal Sentinel-2 data [J]. *Remote Sensing*, 2019, 11(8): 929. DOI:910.3390/rs11080929.
- [16] Srinet R, Nandy S, Padalia H, *et al.* Mapping plant functional types in Northwest Himalayan foothills of India using random forest algorithm in Google Earth Engine [J]. *International Journal of Remote Sensing*, 2020, 41(18): 7296-7309.
- [17] Hijmans R J, Cameron S E, Parra J L, *et al.* Very high resolution interpolated climate surfaces for global land areas [J]. *International Journal of Climatology*, 2005, 25(15) : 1965-1978. DOI:10.1002/joc.1276..
- [18] Main R, Cho M A, Mathieu R, *et al.* An investigation into robust spectral indices for leaf chlorophyll estimation [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2011, 66(6): 751-761.
- [19] Dash J, Curran P. Evaluation of the MERIS terrestrial chlorophyll index (MTCI) [J]. *Advances in Space Research*, 2007,

- 39(1): 100-104.
- [20] Long T, Zhang Z, He G, *et al.* 30 m resolution global annual burned area mapping based on Landsat Images and Google Earth Engine[J]. *Remote Sensing*, 2019, 11(5): 489. DOI: 410.3390/rs11050489.
- [21] Rozenstein O, Haymann N, Kaplan G, *et al.* Validation of the cotton crop coefficient estimation model based on Sentinel-2 imagery and eddy covariance measurements[J]. *Agricultural Water Management*, 2019, 223: 105715. DOI:10.1016/j.agwat.2019.105715.
- [22] Wu C, Niu Z, Tang Q, *et al.* Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation [J]. *Agricultural and forest meteorology*, 2008, 148(8-9): 1230-1241.
- [23] Chandrasekar K, Sessa Sai M, Roy P, *et al.* Land Surface Water Index (LSWI) response to rainfall and NDVI using the MODIS Vegetation Index product[J]. *International Journal of Remote Sensing*, 2010, 31(15): 3987-4005.
- [24] Ahamed T, Tian L, Zhang Y, *et al.* A review of remote sensing methods for biomass feedstock production [J]. *Biomass and bioenergy*, 2011, 35(7): 2455-2469.
- [25] Richardson A D, Duigan S P, Berlyn G P. An evaluation of noninvasive methods to estimate foliar chlorophyll content[J]. *New phytologist*, 2002, 153(1): 185-194.
- [26] Schlerf M, Atzberger C, Hill J. Remote sensing of forest biophysical variables using HyMap imaging spectrometer data[J]. *Remote Sensing of Environment*, 2005, 95(2): 177-194.
- [27] Carlson T N, Ripley D A. On the relation between NDVI, fractional vegetation cover, and leaf area index [J]. *Remote sensing of Environment*, 1997, 62(3): 241-252.
- [28] Nagler P L, Scott R L, Westenburg C, *et al.* Evapotranspiration on western US rivers estimated using the Enhanced Vegetation Index from MODIS and data from eddy covariance and Bowen ratio flux towers[J]. *Remote Sensing of Environment*, 2005, 97(3): 337-351.
- [29] Pu R, Gong P, Yu Q. Comparative analysis of EO-1 ALI and Hyperion, and Landsat ETM+ data for mapping forest crown closure and leaf area index[J]. *Sensors*, 2008, 8(6): 3744-3766.
- [30] Ehammer A, Fritsch S, Conrad C, *et al.* Statistical derivation of fPAR and LAI for irrigated cotton and rice in arid Uzbekistan by combining multi-temporal RapidEye data and ground measurements [C]// *Proceedings of the Remote Sensing for Agriculture, Ecosystems, and Hydrology XII*, F, International Society for Optics and Photonics, 2010.
- [31] Ahmadi K, Kalantar B, Saeidi V, *et al.* Comparison of machine learning methods for mapping the stand characteristics of temperate forests using multi-spectral Sentinel-2 data[J]. *Remote Sensing*, 2020, 12(18): 3019. DOI: 3010.3390/rs12183019.
- [32] Mao W, Lu D, Hou L, *et al.* Comparison of machine-learning methods for urban land-use mapping in Hangzhou City, China [J]. *Remote Sensing*, 2020, 12(17): 2817. DOI: 2810.3390/rs12172817.
- [33] Shao Y, Lunetta R S. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2012, 70: 78-87. DOI:10.3390/rs12172817.
- [34] Clark M L. Comparison of multi-seasonal Landsat 8, Sentinel-2 and hyperspectral images for mapping forest alliances in Northern California [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 159: 26-40.
- [35] Schmidt M, Pringle M, Devadas R, *et al.* A framework for large-area mapping of past and present cropping activity using seasonal Landsat images and time series metrics[J]. *Remote Sensing*, 2016, 8(4): 312. DOI:310.3390/rs8040312.
- [36] Lim J, Kim K M, Kim E H, *et al.* Machine learning for tree species classification using Sentinel-2 spectral information, crown texture, and environmental variables[J]. *Remote Sensing*, 2020, 12(12): 2049. DOI:2010.3390/rs12122049.
- [37] Liu X T, Yuan Q, Ni J. Research advances in modelling plant species distribution in China[J]. *Chinese Journal of Plant Ecology*, 2019, 43(4): 273-283.
- [38] Phillips S J, Anderson R P, Schapire R E. Maximum entropy modeling of species geographic distributions [J]. *Ecological Modelling*, 2006, 190(3-4): 231-259.
- [39] Yi Y J, Cheng X, Yang Z F, *et al.* Maxent modeling for predicting the potential distribution of endangered medicinal plant (*H. riparia* Lour) in Yunnan, China[J]. *Ecological Engineering*, 2016, 92: 260-269.
- [40] Gilani H, Goheer M A, Ahmad H, *et al.* Under predicted climate change: Distribution and ecological niche modelling of six native tree species in Gilgit-Baltistan, Pakistan [J]. *Ecological Indicators*, 2020, 111: 106049.
- [41] Simpson E H. Measurement of diversity [J]. *Nature*, 1949, 163(4148): 688-688.
- [42] Praticò S, Solano F, Di Fazio S, *et al.* Machine learning classification of mediterranean forest habitats in Google Earth Engine based on seasonal Sentinel-2 time-series and input image composition optimisation[J]. *Remote Sensing*, 2021, 13(4): 586. DOI:510.3390/rs13040586.
- [43] Grabska E, Frantz D, Ostapowicz K. Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians[J]. *Remote Sensing of Environment*, 2020, 251: 112103.
- [44] Chandra A, Yao X. Evolving hybrid ensembles of learning machines for better generalisation[J]. *Neurocomputing*, 2006, 69(7-9): 686-700.
- [45] Ranawana R, Palade V. Multi-classifier systems: Review and

- a roadmap for developers[J]. *International journal of hybrid intelligent systems*, 2006, 3(1): 35-61.
- [46] Alonso L, Picos J, Armesto J. Forest cover mapping and pinus species classification using very high-resolution satellite images and random forest [J]. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2021, 3: 203-210.
- [47] Agrillo E, Filipponi F, Pezzarossa A, *et al.* Earth observation and biodiversity big data for forest habitat types classification and mapping[J]. *Remote Sensing*, 2021, 13(7): 1231. DOI: 1210.3390/rs13071231.
- [48] Kollert A, Bremer M, Löw M, *et al.* Exploring the potential of land surface phenology and seasonal cloud free composites of one year of Sentinel-2 imagery for tree species mapping in a mountainous region[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2021, 94: 102208.
- [49] Hill R A, Wilson A K, George M, *et al.* Mapping tree species in temperate deciduous woodland using time-series multi-spectral data[J]. *Applied Vegetation Science*, 2010, 13(1): 86-99. DOI:10.1111/j.1654-1109X.2009.01053.x.
- [50] Naboureh A, Ebrahimi H, Azadbakht M, *et al.* RUESVMs: An ensemble method to handle the class imbalance problem in land cover mapping using Google Earth Engine [J]. *Remote Sensing*, 2020, 12(21):3484. DOI:3410.3390/rs 12213484.
- [51] Yang Y, Xiao P, Cheng Y, *et al.* Ensemble strategy for hard classifying samples in class-imbalanced data set [C] // *Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*, F, 2018.
- [52] Foody G M, Boyd D S, Sanchez-Hernandez C. Mapping a specific class with an ensemble of classifiers[J]. *International Journal of Remote Sensing*, 2007, 28(8): 1733-1746. DOI: 1710.1080/01431160600962566.

Decision Fusion Classification of Forest Dominant Tree Species in Shangri-La Area of Yunnan Province

Fang Panfei¹, Wang Leiguang^{2,3}, Xu Weiheng^{2,3}, Ou Guanglong¹,
Dai Qinling⁴, Li Ruonan¹

(1. College of Forestry, Southwest Forestry University, Kunming 650233, China;

2. Institute of Big Data and Artificial Intelligence, Southwest Forestry University, Kunming 650024, China;

3. Key Laboratory for Forestry and Ecological Big Data of National Forestry and Grassland Administration, Southwest Forestry University, Kunming 650024, China;

4. Art and Design College, Southwest Forestry University, Kunming 650233, China)

Abstract: Based on Google Earth Engine (GEE) cloud computing platform, we collaborate with Sentinel-2 images, WordClim bioclimatic data, SRTM topographic data, forest resources planning and design survey data and other data, and use Random Forest (RF), Support Vector Machine (SVM) and Maximum Entropy (MaxEnt) machine learning algorithms were used as component classifiers to carry out the study of dominant tree species classification with multi-source features and multi-classifier decision fusion. Two serially integrated and three Bayesian parallel integrated models were constructed by the three component classifiers for determining the spatial distribution of 10 major dominant tree species in Shangri-La region of Yunnan. The classification results showed that the overall accuracy of the three component classifiers was lower than 67.17%, the overall accuracy of the three parallel integration methods was comparable, about 72%, the accuracy of the two serial integration methods was higher than 78.48%. Among them, the MaxEnt SVM serial integration method obtained the best accuracy (OA: 80.66%, Kappa: 0.78), which improved the accuracy compared with the component classifiers by at least 13.49%. The study shows that the decision fusion method has higher accuracy than the component classifier in dominant tree species classification and effectively improves the classification accuracy of small sample tree species, which can be used for dominant tree species classification in large mountainous areas.

Key words: Dominant tree species; Machine learning; Decision fusion; GEE