



基于人工智能算法的 CYGNSS 数据土壤水分反演

摘要

利用 CYGNSS 数据估计地表土壤水分(SM)近年来获得了极大的关注,但效率和精度有待进一步提升.本文提出了一种预分类策略,结合人工智能算法(AI),利用 CYGNSS 数据预测土壤水分.此策略能够在人工智能算法的基础上进一步提高土壤水分预测的精确度,具有较好的普适性和易用性.本文使用了 2018 年全年中国地区的实地土壤水分数据作为地面真实参考数据进行建模以及预测.结果证实预测土壤水分与参考真实数据具有良好的一致性.基于 CYGNSS 数据预测的土壤水分与实地土壤水分参考数据比对,其相关系数高达 0.8,平均均方根误差(RMSE)和平均无偏均方根误差(ubRMSE)分别为 $0.059 \text{ cm}^3/\text{cm}^3$ 和 $0.050 \text{ cm}^3/\text{cm}^3$.研究表明,预分类策略的人工智能算法可明显提高 CYGNSS 预测土壤水分的精确度,其简单操作性也使其可广泛应用于其他回归和预测研究领域,具有较好的泛化性和拓展性.

关键词

全球导航卫星系统反射测量法;土壤水分;CYGNSS;人工智能

中图分类号 S152.7;S127

文献标志码 A

收稿日期 2021-10-08

资助项目 国家自然科学基金(42001375,42001362);江苏省自然科学基金(BK20180765)

作者简介

贾燕,女,博士,讲师,研究方向为 GNSS-R 遥感技术与应用.jiayan@njupt.edu.cn

- 1 南京邮电大学 地理与生物信息学院,南京,210023
- 2 南京信息工程大学 遥感与测绘工程学院,南京,210044
- 3 南京信息工程大学 江苏省协同精密导航定位与智能应用工程研究中心,南京,210044
- 4 中国科学院上海天文台,上海,200030

0 引言

地表土壤水分是地表与大气相互作用的重要因子,是作物生长、发育的基本条件.它不仅对陆地表面蒸散、水的运移、碳循环有很强的调控作用,而且也是气候、水文、生态、农业等领域衡量土壤干旱程度的重要指标^[1-3],因此,及时准确地获取地表土壤水分信息具有重要的意义.但是,由于监测设备的高成本以及点观测的局限性(小范围观测),传统的现场观测无法满足持续监测大面积土壤水分含量的需求^[4].微波遥感技术的兴起和发展使得传统的点测量转移到面测量,为获得区域和全球尺度的高分辨率的土壤水分信息提供了可能.

目前,许多被动微波卫星或传感器已被用于观测地表土壤湿度(<5 cm),如美国宇航局的 AMSR-E(先进微波扫描辐射计-地球观测系统)^[5]、欧洲航天局的 SMAP(土壤湿度被动和主动)^[6]和 SMOS(土壤湿度和海洋盐度)^[7].尽管使用微波传感器可以获得高精度的土壤水分产品,例如,SMAP 36 km 土壤水分产品的误差约为 $0.04 \text{ m}^3/\text{m}^3$ ^[8],但其 2~3 d 的较长重访周期限制了其更高的时间分辨率(1 d)应用.

GNSS-Reflectometry (GNSS-R)反射信号遥感是一种介于主动和被动遥感之间的新型遥感探测方法,可以看作是一个非合作人工辐射源、收发分置、多发单收的多基地 L 波段雷达系统,从而兼具主动和被动遥感两者的优点^[9-10]:1)它利用全球卫星导航系统反射信号进行测量,不需要额外的发射机,这使得开发轻便、紧凑和低成本接收机系统得以实现;2)系统配置灵活,可自行设计接收多种高度、多角度和多极化的目标物反射信号,为多种接收方式、算法研究提供了可能.随着 GNSS-R 反演土壤水分研究的不断深入,具有长时间序列观测数据的新星座观测计划成为 GNSS-R 土壤水分反演的新途径.最新的 CYGNSS 星座系统,含有 8 颗微小卫星,可同时接收来自 GPS 卫星的直射信号和地面反射信号^[9],大大增加了地球可观测面积.它的重访周期仅为 4 h,空间分辨率最高可达到 $0.5 \text{ km}^{[11]}$.CYGNSS 星座系统提供的高精度优良数据,为实现全球高时空分辨率的土壤水分长期动态监测提供了有力的支撑^[12-24].

Eroglu 等^[17]提出了一种基于人工神经网络(ANN)的日土壤水分反演方法,其设计的 ANN 输入部分包括 CYGNSS 数据和其他辅助数据:归一化差异植被指数(NDVI)、植被含水量(VWC)、地形高程、地

形坡度和表面粗糙度.反演结果的 ubRMSE (无偏均方根误差)为 $0.0544 \text{ cm}^3/\text{cm}^3$, R 为 0.90,取得了较好的研究成果.虽然 ANN 能够很好地反演土壤水分,但该方法十分依赖训练数据集.倘若训练数据集不能代表所有实测数据,就会出现反演值溢出的情况.Yang 等^[19]采用了类似的算法评估了星载 GNSS-R 星座系统 (TDS 和 CYGNSS) SM 预测性能. R 为 0.79, ubRMSE 为 $0.062 \text{ cm}^3/\text{cm}^3$,并采用了 6 个辅助参数.值得注意的是,目前基于人工智能的 CYGNSS 数据反演土壤水分研究^[17-19],大多使用 SMAP 作为参考和验证数据,且依赖于大量的辅助输入参数,例如海拔高度、地形坡度等.

因此,本文提出一种基于预分类的人工智能算法,利用 CYGNSS 数据进行 SM 预测和估计.采用上述方案,其普适性和实用性较强且需要的辅助变量较少,学习模型的复杂度低,训练效率高且训练过程简易,同时预测结果的精度高.为 CYGNSS SM 估计提供了新的思路 and 手段.

1 数据描述

本章主要描述 CYGNSS 数据获取与预处理过程,并介绍了 SMAP 数据以及实地土壤水分参考数据及来源.

1.1 CYGNSS 数据

本文拟采用的土壤水分数据产品主要包括星载 CYGNSS 以及实测土壤水分数据.其中,星载 CYGNSS 有 8 颗微卫星,每颗卫星每秒可以同时提供 4 个地面反射测量值(免费获取:<https://podaac.jpl.nasa.gov>)^[11],因此每秒可以同时获取总共 32 个观测数据.CYGNSS 原始数据包括每个镜面反射点 (SP) 处的双地基雷达横截面值 (BRCS) 和信噪比 (SNR),以及相关的几何测量和导航定位信息,例如入射角、SP 坐标、从 SP 到发射端和接收端的距离等.本文中使用的 2018 年整年的 CYGNSS 数据用于建模和预测,并把每日 CYGNSS 数据样本进行重投影、重采样到 EASE-Grid (该网格为等积圆柱投影) 网格上,便于后续的计算与比较验证.

本文拟采用中国区域范围内的 CYGNSS 数据中 SP 处 SNR 超过 0 dB 的数据进行土壤水分反演,以确保数据的可靠性.同时,保留 BRCS 峰值在时延轴第 4 与第 15 位的数据,天线增益大于零的数据以及仰角大于 30° 的数据,相关类似操作也在文献 [17-20] 中采用.另外,由于 CYGNSS 未直接提供 GNSS-R

双基雷达反演中所需的地表反射率,因此需要通过 CYGNSS 已提供的相关数据来推算地表反射率,将在下一节进行详细介绍.

1.2 SMAP 数据

SMAP 提供日土壤水分产品,且平均每隔 3 d 左右,其数据可覆盖到全球 $\pm 45^\circ$ 纬度内的陆地区域.本文拟采用 SMAP L3 级产品, EASE-Grid 版本 6 提供的日土壤水分产品,该数据的空间分辨率为 $36 \text{ km} \times 36 \text{ km}$ (免费获取:<https://nsidc.org/data/SPL3SMP/>).SMAP 的数据包含土壤水分估计值、质量标记 (用于过滤 SMAP 数据)、粗糙度和植被不透明度 (vod, 或 τ) 以及相关的轨道坐标值^[25].本文主要使用 SMAP 中的粗糙度和植被不透明度 τ 作为辅助变量参与建模,并使用质量标记“反演成功”作为数据筛选的条件.因为单日的 SMAP 数据覆盖率较低,我们将 SMAP 数据都进行了 3 d 平均的操作便于其参与后续的土壤水分建模运算,并将 SMAP 的粗糙度和 τ 数据映射到 CYGNSS 数据使用的 EASE-Grid 网格上.

1.3 参考数据

将上述数据作为输入变量,建立土壤水分预测模型.输出的土壤水分参考数据将采用中国土壤水分自动观测站收集的实地地表 SM 数据.该观测网由多个站点组成,本文采用的站点分布及其覆盖的区域如图 1 所示.每个站点提供每小时土壤表面以下 0 至 100 cm 深度的 SM 测量,间隔 10 cm.根据不同的土壤含水量和土壤类型,GNSS-R 信号在土壤中的穿透深度可以从几厘米到大约 20 厘米不等^[15].因此采用地表上层 10 cm 的实地观测 SM 数据,视为与 GNSS-R 技术反演土壤水分匹配的最佳值.在本研究中,每天逐小时的地表土壤水分数据被用来作为参考真实土壤水分数据.此外,还提供了采样地点的地理位置信息 (包括纬度、经度).

2 估计方法

本章详细叙述了从 CYGNSS 数据计算反射率以及土壤水分的建模和预测过程,包括地表反射率的计算、几种用于土壤水分反演的经典人工智能算法,以及预分类人工智能算法的模型构建和反演 SM 的过程.

2.1 CYGNSS 观测量计算

拟利用处理好的 CYGNSS 数据,在顾及地表粗

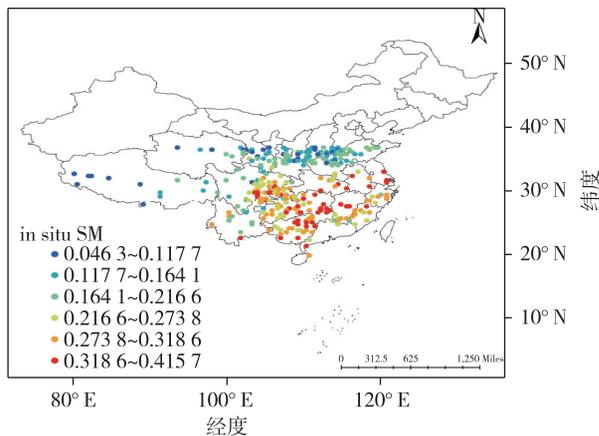


图1 本研究中国土壤水分自动观测站的分布
Fig. 1 Distributions of automatic soil moisture observation stations in China used for this study

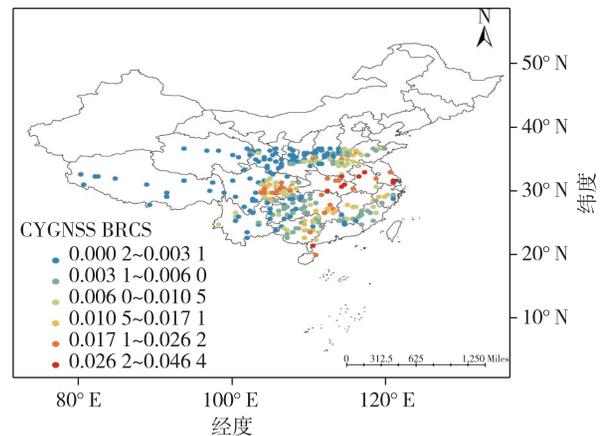


图2 用于实地土壤水分反演的CYGNSS (2018-01-01) 反射率分布
Fig. 2 An example of CYGNSS reflectivity samples for SM estimation in China on Jan.1st, 2018

糙度和植被的情况下,进行土壤水分反演.对于有植被覆盖的区域,地表的反射率^[12-25]可表示为以下的形式:

$$\Gamma(\theta) = R(\theta)^2 \gamma^2 \exp(-4k^2 s^2 \cos(\theta)), \quad (1)$$

其中,入射角为 θ , R 为非涅耳反射系数,透射率 γ 包含了植被对信号传播的衰减程度,指数项表示地表粗糙度效应,其中 k 为信号波数, s 为地表均方根高度.

反射率 Γ 和土壤水分呈线性正相关的关系,因此 Γ 可被视为与土壤水分相关性最大的特征变量.另外,由于透射率 γ 是植被不透明度 τ 的函数,其形式为 $\gamma = \exp(-\tau \sec \theta)$,所以拟将SM视作 Γ 、 τ 和地表粗糙度3个变量的函数.本文拟采用SMAP数据中的植被不透明度 τ 和地表粗糙度作为辅助数据,地表反射率可由CYGNSS数据获取.

根据光滑地表相干反射分量的理论假设,并考虑CYGNSS提供的BRCS值,地表的反射率可由下列公式^[17]求得:

$$\Gamma_{brcs} = \frac{\sigma(R_t + R_r)^2}{4\pi(R_t R_r)^2}, \quad (2)$$

其中 R_t 和 R_r 分别表示发射机和接收机到SP的距离,这些相关参数由CYGNSS数据提供.值得注意的是,反射率的求解可采用多种途径,由于采用BRCS获得的 Γ_{brcs} 在土壤水分反演中已被证实取得的效果最佳^[17],故本文采用式(2)的方法求解反射率,结果如图2所示.

2.2 人工智能算法用于土壤水分反演

机器学习(ML)是一种主流的实现人工智能的

方法,而机器学习最基本的做法,是使用算法来解析数据、从中学习,然后对真实世界中的事件做出决策和预测.与传统的为解决特定任务、硬编码的软件程序不同,机器学习用大量的数据来“训练”,通过各种算法从数据中学习如何完成任务.机器学习历经70年的曲折发展,在很多方面收获了突破性进展,经典的机器学习的研究方向主要包括决策树、随机森林、人工神经网络、SVM等,且已经被运用到了土壤水分的反演研究中.

随机森林^[26]利用bootstrap重抽样方法从数据集中抽取多个样本,对每个样本分配一个决策树进行决策,即建立多棵决策树,形成一个决策树“森林”,然后将多棵决策树进行组合,最后通过投票方法得到最终预测结果.该方法能够有效地提高对新样本的分类准确率,降低模型的过拟合能力.

人工神经网络(Artificial Neural Network)模型,简称ANN,是1980年代以来人工智能领域兴起的研究热点之一^[17].从生物学的角度来说,它是模拟人脑对信息处理的过程,是对人脑神经系统的数学模型抽象.神经网络实现一种数学运算,由大量的神经元相互连接而成,每个神经元可以看作是一种运算输出函数,该模型可以用于拟合复杂的数学函数.人工神经网络模型分为多层前向神经网络、自组织神经网络、Hopfield神经网络等.

XGBoost算法,全称Extreme Gradient Boosting,出现于Chen等^[27]2016年发表的论文中,在Kaggle等比赛中该算法取得佳绩.XGBoost以决策树模型为基础,是对梯度提升树的改进算法.该算法是一种适合处理

稀疏数据的树学习算法,支持并行运算和分布式运算,通过对损失函数进行二阶泰勒展开来达到二次优化的目的,因此该算法学习效果好、分类精度高、处理速度快,并具有强大的可伸缩性,被人们广泛应用于包括股票选择策略、电力系统预测等诸多领域。

2.3 基于预分类人工智能算法的土壤水分预测模型

GNSS-R 中接收的地表反射信号主要成分为地表的相干反射分量。随着地表粗糙度和植被的增加,非相干散射分量增加,同时相干分量降低。通过对反射率、粗糙度和植被这 3 个主要参数的拟合计算,可以获得 SM 估计值^[16]。因此在本文的 SM 预测模型中,CYGNSS 反射率作为主要变量,SMAP 粗糙度系数和植被不透明度作为辅助变量。

之前的研究^[17-19]试图添加各种各样的辅助数据,以提高 SM 估计的准确性。可以发现,添加的大部分辅助数据都与地形有关,如地形坡度和土壤质地^[17-19]。这些辅助数据已表明其有能力提高预测的准确性,但辅助数据量较大且种类较多,尤其是在全球尺度下,获取稳定的多种辅助源变得较为困难。因此,本文提出了一种新的预分类策略,该策略基于传统的 ML 回归方法,采用分类建模的思想,来最小化不同土地类型数据对学习模型的影响,以简单实用的方式提高 SM 估计精度。图 3 展示了通过使用 ML 回归和预分类策略方法对 CYGNSS 反演 SM 进行建模学习和回归预测的流程。

本文使用 2018 年的 CYGNSS、SMAP 和相应的实地参考土壤水分数据建立学习模型,模型的输入参数包括 CYGNSS 反射率,并将其视为模型的主要输入参数(图 3)。SMAP 的粗糙度系数(rou)和植被不透明度(vod)用作模型的辅助参数,实地 SM 数据被用作训练和验证所提出的 ML 方法的真实参考数据。根据 SMAP 提供的国际地圈-生物圈计划(IGBP)土地类型标识数据,对不同土地类型的样本进行分类重组,然

后分别使用同类样本建立若干子模型(例如,土地类型 1、2 等)进行 SM 建模以及预测。

将提出的预分类 ML 算法结果与传统 ML 回归进行比较,并采用 10 倍交叉验证(CV)训练和验证所提出的预分类模型回归预测的可行性,根据均方根误差指标(RMSE)结果选择最佳算法。10 倍交叉验证是常用且流行的方法,与其他方法相比,它通常会产生较少的偏差。因为它确保了来自原始数据集的每个数据都有机会出现在训练和测试集中。将整个数据集随机分成 10 个部分(取决于数据大小),然后用 9 份作为训练集对模型进行拟合,并使用剩余的集合对模型进行验证。记下 RMSE 作为性能指标。重复此过程,直到第 10 份用作测试集。将每次迭代中记录的 RMSE 的平均值作为最终评估性能指标。

3 结果与评估

本章对提出的预分类机器学习反演算法进行了建模以及验证。本文选取了 2018 年的 CYGNSS 数据及其对应的实地真实数据。基于第 2 章描述的土壤水分预测模型结构,对基于预分类的各种人工智能算法结果进行分析及验证。

3.1 多种人工智能算法反演土壤水分结果比较

本文采用了不同的传统 ML(RF、SVM、XGBoost)和高级神经网络(ANN)算法进行土壤水分预测的计算,以及展示了各个算法在增加了预分类策略后的效果比较(表 1)。本文中人工智能算法的选择基于两个原因:其一是此算法已经被证实可用于土壤水分反演并取得了较好的预测效果,其二是此算法较为经典,运用广泛且执行效率较高。根据 SMAP 提供的 IGBP 土地分类,CYGNSS 数据在中国区域共包含 7 个类别。通过使用 2018 年的数据分类建立预测模型,展示了所提出的预分类 ML 回归在 CYGNSS 土壤水分预测中的性能。

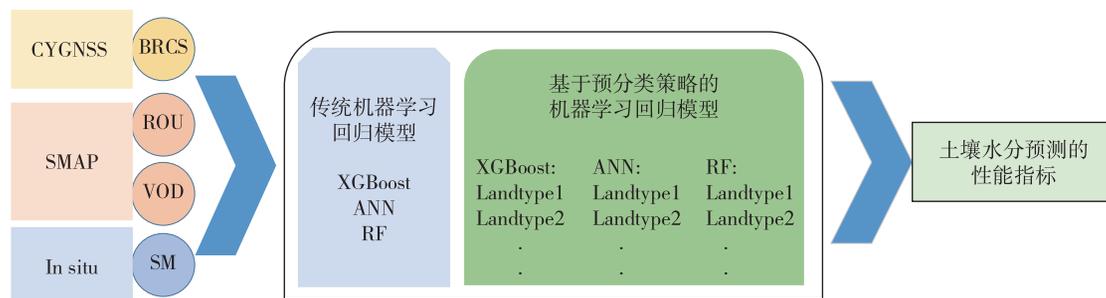


图 3 基于预分类 ML 算法的土壤水分预测模型结构示意图

Fig. 3 An example diagram of SM prediction model based on pre-classification ML algorithm

表1 CYGNSS SM 估计的性能统计

Table 1 Performance statistics of CYGNSS estimation

cm³/cm³

IGBP 分类	RMSE					
	XGBoost		RF		ANN	
	预分类	传统方法	预分类	传统方法	预分类	传统方法
常绿针叶林	0.077	0.083	0.082	0.087	0.083	0.087
灌丛	0.056	0.068	0.060	0.061	0.067	0.071
高山亚高山草地	0.061	0.068	0.061	0.067	0.067	0.070
平原草地	0.055	0.067	0.060	0.062	0.064	0.070
草原	0.058	0.067	0.064	0.067	0.066	0.068
河流	0.059	0.071	0.062	0.068	0.069	0.073
沼泽	0.032	0.039	0.037	0.038	0.039	0.040
总结果	0.062	0.070	0.066	0.071	0.069	0.073

与传统的回归模型比较,SM 预测效果较好,提出的预分类(子模型建模)策略方法在所有算法(传统 ML 和 DL)以及所有土地类型中均取得了良好的效果,使用预分类策略时 RMSE 有了明显下降.此外,通过比较不同的 ML 算法,RF 的性能优于 ANN,这与文献[18,20]一致.此外,XGBoost 的性能最好, RMSE 最小为 0.062 cm³/cm³.在后续的分析中,采用预分类策略的 XGBoost 作为最优预测模型,展示 CYGNSS SM 估计在土壤水分预测和时空分布方面的表现.

在图 4 中,给出了反演 SM 数据测试集密度的示例来进一步展示 CYGNSS 模型预测的 SM 和实地参考 SM 数据之间的对应关系.密度图显示了使用 XGBoost 预分类策略进行 SM 估计的结果,其与实地 SM 值有相当好的一致性,总体相关性(*R*)为 0.8.尤其是当数据最密集时,点的分布沿 1:1 线居中.但是,如果数据密度较低,则稍稍显示偏离直线的趋势.斜率呈下降趋势,小于 1,这意味着 CYGNSS 预测模型在某种程度上低估了 SM 值.这一现象在之前的研究^[16]中也有报道,原因可能是较高土壤水分的地表通常植被生长较为茂密,植被含水量也较高,导致多种非相干分量增加、相干分量减少,此时反射率的变化不足以完整和正确地表达地表湿度的变化,而且土壤湿度和信号相干分量具有正相关的关系,导致学习模型不能正确提取高土壤湿度地表的特征,因此模型预测的 SM 较低.此现象会在未来的工作中进行进一步验证和分析.

3.2 XGBoost 预分类的 CYGNSS SM 反演分布结果

本节采用性能最优的预分类 XGBoost 算法进行土壤水分预测的分析,并展示了其在全国范围内的实地测量参考数据站点中预测土壤水分的分布及其

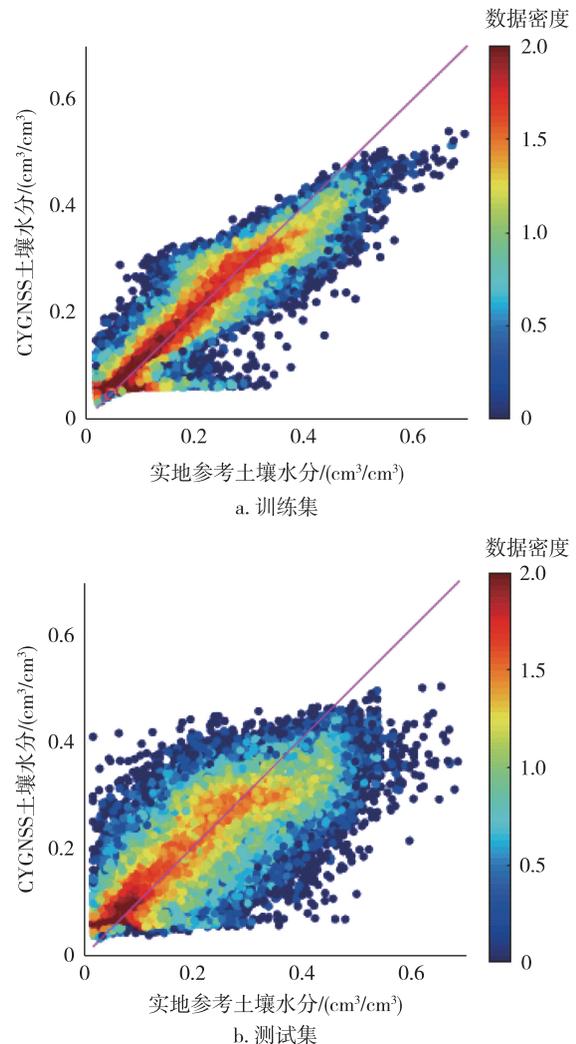


图4 CYGNSS SM 反演结果在训练集(a)和测试集(b)的密度
Fig. 4 Density plots of CYGNSS SM retrievals using training (a) and test (b) data

性能.其中表 2 展示了 CYGNSS SM 预测土壤水分和实地土壤水分观测站 SM 土壤水分值的性能指标

值,把全国土壤水分站按照地表特征以及地域行政区域划分为 5 个土壤水分网络.针对 5 个 SM 网络中的真实地表土壤水分,分别展示其总体和区域的 RMSE 和无偏 RMSE (ubRMSE) 等性能指标,有助于与其他研究进行对比和比较^[17-19].

表 2 基于预分类的 XGBoost 算法相对于现场真实测量 SM 的总体性能

Table 2 Overall performance of the predicted SM against in situ measurements cm^3/cm^3

实地参考数据	平均值 ubRMSE	平均值 RMSE
所有数据点	0.050	0.059
贵州地区	0.057	0.066
中东部地区	0.046	0.055
东南沿海地区	0.042	0.051
西北地区	0.048	0.057
其他数据点	0.040	0.053

在表 2 中,基于预分类的 XGBoost CYGNSS 模型反演土壤水分预测结果相对于真实参考土壤水分达到了较高的精确度,总体平均为 ubRMSE ($0.050 \text{ cm}^3/\text{cm}^3$) 和 RMSE ($0.059 \text{ cm}^3/\text{cm}^3$).此外,不同的区域主要的土地类型不同,因此预测的土壤水分结果有些许差异.特别是,贵州地区相对于其他区域,其土壤水分预测误差较大, RMSE 为 $0.066 \text{ cm}^3/\text{cm}^3$, ubRMSE 为 $0.057 \text{ cm}^3/\text{cm}^3$.原因可能由于此地区所含有的主要地表覆盖是树木和森林,且有大量高低起伏的山体.因此,密集的植被和高地特征显著影响反射信号中相干分量的接收,这一现象在上一节中也进行了详细的阐述.

图 5 展示了所有采用的现场观测站点及其各自土壤水分预测精度的分布.这里我们分别计算了每日平均 CYGNSS 土壤水分估计和真实参考数据之间的精度指标 RMSE.如上所述, RMSE 值的波动受到土壤质地和周围环境的影响.一般来说,基于 CYGNSS 预测的 SM 与现场观测的真实土壤水分在绝大多数的站点都表现出了较高的精确度,进一步表明预分类人工智能 CYGNSS 方法得到的 SM 精确度较高,可用于估计预期的 SM 值.

4 结束语

本文提出了一种基于预分类的人工智能 CYGNSS 数据反演土壤水分的方法,以求用较少的辅助变量达到较好的土壤水分预测效果.新的预分类策略具有高集成度的特点,根据不同的土地类型

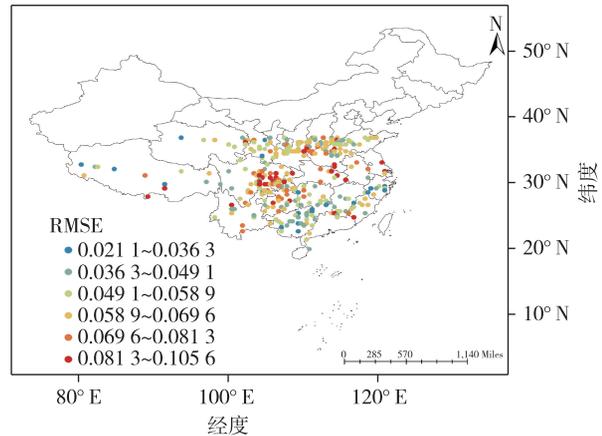


图 5 CYGNSS SM 各站点土壤水分反演结果分布

Fig. 5 Accuracy distribution of soil moisture retrieved from CYGNSS

对数据进行预分类,然后分别对每个类别的数据进行建模学习,构建土壤水分预测子模型.对具有相似特点的同类型土壤质地的数据集中建模,这一操作有助于最小化不同土地类型给土壤水分预测带来的影响,能够帮助机器学习模型更准确地发现和提取数据特征,进而建立更精确的学习模型,从而提高土壤水分预测值.输入变量为 CYGNSS 反射率、SMAP 植被不透明度和粗糙度,输出变量为实地参考土壤水分值.采用预分类策略的土壤水分反演结果与传统的 ML 回归方法进行了比较,反演结果均展现了高准确度.预分类策略显示了更强的预测能力,多种典型 ML 方法在采用预分类策略时其 RMSE 明显下降.其中,预分类的 XGBoost 算法预测效果最佳, RMSE 为 $0.062 \text{ cm}^3/\text{cm}^3$, 相关系数 $R=0.8$.此外,分类结果显示不同的地表种类对土壤水分预测精度有影响.值得注意的是,我们发现 CYGNSS 预测模型在某种程度上低估了 SM 值.推测原因可能是由于高土壤水分地区通常植被较为茂密,植被含水量也较高,非相干分量增加,导致接收的反射信号相干分量不能够准确表达真实地表土壤湿度的情况,进而致使模型预测的 SM 较低.此现象会在未来的工作中进行进一步验证和分析.此外,本文采用了实地真实土壤水分参与建模,其数据量较少,今后可采用更大尺度的数据进行建模,预计能更有利于预分类人工智能算法发挥作用.本文采用预分类的人工智能算法反演土壤 SM 方案,普适性和实用性较强且应用的辅助变量少,学习模型的复杂度低,训练效率高且训练过程简易,同时预测结果的精度高,其通用性较强也可用于其他的回归预测应用研究中.

参考文献

References

- [1] Dobriyal P, Qureshi A, Badola R, et al. A review of the methods available for estimating soil moisture and its implications for water resource management [J]. *Journal of Hydrology*, 2012, 458/459: 110-117
- [2] Ban W, Yu K G, Zhang X H. GEO-satellite-based reflectometry for soil moisture estimation: signal modeling and algorithm development [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(3): 1829-1838
- [3] Dobson M C, Ulaby F T, Hallikainen M T, et al. Microwave dielectric behavior of wet soil-part II: dielectric mixing models [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 1985, GE-23(1): 35-46
- [4] Lu C X, Feng G L, Zheng Y X, et al. Real-time retrieval of precipitable water vapor from Galileo observations by using the MGEX network [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(7): 4743-4753
- [5] Njoku E G, Jackson T J, Lakshmi V, et al. Soil moisture retrieval from AMSR-E [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2003, 41(2): 215-229
- [6] Entekhabi D, Njoku E G, O'Neill P E, et al. The soil moisture active passive (SMAP) mission [J]. *Proceedings of the IEEE*, 2010, 98(5): 704-716
- [7] Kerr Y H, Waldteufel P, Wigneron J P, et al. Soil moisture retrieval from space: the soil moisture and ocean salinity (SMOS) mission [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2001, 39(8): 1729-1735
- [8] Chan S K, Bindlish R, O'Neill P E, et al. Assessment of the SMAP passive soil moisture product [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(8): 4994-5007
- [9] 金双根, 张勤耘, 钱晓东. 全球导航卫星系统反射测量 (GNSS+R) 最新进展与应用前景 [J]. *测绘学报*, 2017, 46(10): 1389-1398
JIN Shuanggen, ZHANG Qinyun, QIAN Xiaodong. New progress and application prospects of global navigation satellite system reflectometry (GNSS+R) [J]. *Acta Geodaetica et Cartographica Sinica*, 2017, 46(10): 1389-1398
- [10] 李黄, 夏青, 尹聪, 等. 我国 GNSS-R 遥感技术的研究现状与未来发展趋势 [J]. *雷达学报*, 2013, 2(4): 389-399
LI Huang, XIA Qing, YIN Cong, et al. The current status of research on GNSS-R remote sensing technology in China and future development [J]. *Journal of Radars*, 2013, 2(4): 389-399
- [11] Ruf C S, Atlas R, Chang P S, et al. New ocean winds satellite mission to probe hurricanes and tropical convection [J]. *Bulletin of the American Meteorological Society*, 2016, 97(3): 385-395
- [12] Chew C C, Small E E. Soil moisture sensing using spaceborne GNSS reflections: comparison of CYGNSS reflectivity to SMAP soil moisture [J]. *Geophysical Research Letters*, 2018, 45(9): 4049-4057
- [13] Kim H, Lakshmi V. Use of cyclone global navigation satellite system (CyGNSS) observations for estimation of soil moisture [J]. *Geophysical Research Letters*, 2018, 45(16): 8272-8282
- [14] Al-Khaldi M M, Johnson J T, O'Brien A J, et al. Time-series retrieval of soil moisture using CYGNSS [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(7): 4322-4331
- [15] Yan Q Y, Gong S Q, Jin S G, et al. Near real-time soil moisture in China retrieved from CyGNSS reflectivity [J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 99: 1-5
- [16] Clarizia M P, Pierdicca N, Costantini F, et al. Analysis of CYGNSS data for soil moisture retrieval [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019, 12(7): 2227-2235
- [17] Eroglu O, Kurum M, Boyd D, et al. High spatio-temporal resolution CYGNSS soil moisture estimates using artificial neural networks [J]. *Remote Sensing*, 2019, 11(19): 2272. DOI: 10.3390/rs11192272
- [18] Senyurek V, Lei F N, Boyd D, et al. Machine learning-based CYGNSS soil moisture estimates over ISMN sites in CONUS [J]. *Remote Sensing*, 2020, 12(7): 1168. DOI: 10.3390/rs12071168
- [19] Yang T, Wan W, Sun Z G, et al. Comprehensive evaluation of using TechDemoSat-1 and CYGNSS data to estimate soil moisture over mainland China [J]. *Remote Sensing*, 2020, 12(11): 1699. DOI: 10.3390/rs12111699
- [20] Jia Y, Jin S G, Chen H L, et al. Temporal-spatial soil moisture estimation from CYGNSS using machine learning regression with a preclassification approach [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 4879-4893
- [21] Jia Y, Jin S G, Savi P, et al. GNSS-R soil moisture retrieval based on a XGboost machine learning aided method: performance and validation [J]. *Remote Sensing*, 2019, 11(14): 1655. DOI: 10.3390/rs11141655
- [22] Santi E, Pettinato S, Paloscia S, et al. Soil moisture and forest biomass retrieval on a global scale by using CyGNSS data and artificial neural networks [C] // IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. September 26 - October 2, 2020, Waikoloa, HI, USA. IEEE, 2020: 5905-5908
- [23] Yan Q Y, Huang W M, Jin S G, et al. Pan-tropical soil moisture mapping based on a three-layer model from CYGNSS GNSS-R data [J]. *Remote Sensing of Environment*, 2020, 247: 111944. DOI: 10.1016/j.rse.2020.111944
- [24] Calabria A, Molina I, Jin S G. Soil moisture content from GNSS reflectometry using dielectric permittivity from Fresnel reflection coefficients [J]. *Remote Sensing*, 2020, 12(1): 122. DOI: 10.3390/rs12010122
- [25] NASA Distributed Active Archive Center (DAAC) at National Snow and Ice Data Center (NSIDC). SMAP L3 radiometer global daily 36 km EASE-grid soil moisture [EB/OL]. [2021-10-01]. <https://nsidc.org/data/SPL3SMP/versions/7>
- [26] Breiman L. Random forests [J]. *Machine Learning*, 2001, 45(1): 5-32

[27] Chen T Q, Guestrin C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data

Mining. 13 – 17 August 2016. San Francisco, CA, USA. 2016:785-794

Retrievals of soil moisture from the CYGNSS data based on artificial intelligence algorithms

JIA Yan¹ JIN Shuanggen^{2,3,4} YAN Qingyun^{2,3} GUO Xiantao¹

1 School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing 210023

2 School of Remote Sensing & Geomatics Engineering, Nanjing University of Information Science & Technology, Nanjing 210044

3 Jiangsu Engineering Center for Collaborative Navigation/Positioning and Smart Applications,
Nanjing University of Information Science & Technology, Nanjing 210044

4 Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030

Abstract Retrieving surface Soil Moisture (SM) from CYGNSS has attracted great attention in recent years, yet its accuracy and efficiency should be further improved. Here, a pre-classification strategy combined with Artificial Intelligence (AI) algorithm is proposed to predict SM from CYGNSS data. This strategy can improve the accuracy of SM estimation due to the use of AI algorithm and is versatile and easy to use. The field SM data of China in 2018 are used as real ground truth values for modeling and prediction. The results show that the predicted SM is in good agreement with the referenced SM. The correlation coefficient (R) between SM retrieved from CYGNSS and ground truth data is as high as 0.8, and the mean values of Root Mean Square Error (RMSE) and unbiased root mean square error (ubRMSE) are $0.059 \text{ cm}^3/\text{cm}^3$ and $0.050 \text{ cm}^3/\text{cm}^3$, respectively. Meanwhile, the results show that the AI-based pre-classification strategy not only significantly improves the accuracy of SM estimation from CYGNSS, but is applicable to other regression and prediction fields for its good generalization and expansibility.

Key words global navigation satellite system-reflectometry (GNSS-R); soil moisture (SM); cyclone global navigation satellite system (CYGNSS); artificial intelligence (AI)