

李卓群

男 | 15004059308 | m15004059308@163.com | 中国科学院大学 | 气象学专业 | 硕士研究生
GitHub: <https://github.com/lizhuoq>

教育背景

中国科学院大学, 气象学, 硕士

2022.9 - 2025.7

- 主修课程: 机器学习 (94), 深度学习 (87), 高级人工智能 (85), 算法中的最优化方法 (85), 矩阵分析与应用 (82), Python 空间数据处理 (82)
- 荣誉奖项: Kaggle 竞赛金牌和银牌各一枚, 2024CCF 大数据与计算智能大赛第 8 名
- 研究方向: 机器学习在气象学的应用
- 发表论文: 一篇 SCI 期刊 (JCR 一区), 一篇中国科技核心

成都信息工程大学, 大气科学, 本科

2018.9 - 2022.7

- 主修课程: 动力气象学 (90), 数值计算方法 (90), 数值天气预报 (85), 天气学原理 (80)
- 荣誉奖项: 校三等奖学金一次, 校学业优良奖两次
- 发表论文: 一篇中文核心

项目经历

基于机器学习的青藏高原高分辨率表层土壤湿度数据的建立

2023.6 - 2023.9

- 青藏高原土壤湿度数据对生态与气候研究至关重要, 但现有数据产品存在分辨率低或误差大的问题, 亟需高精度数据。
- 基于多源数据 (遥感、气象、土壤属性等), 构建机器学习模型实现高分辨率 (500m) 土壤湿度反演, 并验证其优于主流数据产品 (ERA5-Land、ESA CCI SM 等)。
- 采用梯度增强模型 (XGBoost、CatBoost、LightGBM) 和深度学习模型 (LSTM、Transformer) 对比实验, 通过 5 次 5 折交叉验证评估性能, 结合特征重要性分析优化模型推理过程。
- XGBoost 表现最优 ($R=0.783$, $RMSE=0.0605$), 集成模型显著降低误差; 反演结果分辨率与精度均优于现有产品, 成果支撑高原生态水文研究。

基于对抗生成网络对三江源 CMIP6 土壤湿度数据统计校正

2023.10 - 2024.3

- 三江源区是生态敏感区, 但缺乏气候变化情景下的高分辨率土壤湿度数据。现有的 CMIP6 模型在这种复杂地形下分辨率较低, 性能不稳定。
- 开发基于 Gan 的框架, 将 CMIP6 数据融合并降尺度到高分辨率 (0.1°), 然后分析其在四种 SSP 情景 (1950-2100) 下的时空演变。
- 设计了 UNet-GAN 模型 (UNet 作为生成器), 将 CMIP6 与高分辨率历史数据集成, 实现融合和降尺度; 将 Bagging 应用于五个交叉验证的模型, 以进行稳健的预测。使用统计指标和显著性测试分析深度、季节和时期的趋势。
- 生成的 $CMIP6_{UNet-Gan}$ 数据集 ($R^2=0.819$) 显示, 到 2100 年, 土壤湿度显著增加 (表层增加 5.5 - 11.5%), 其中春冬季和长江西北部地区的趋势最强。结果发表在 Remote Sensing 上。

LEAP - Atmospheric Physics using AI (ClimSim) 竞赛, 团队 (队长)

2024.4 - 2024.7

- 该比赛由 Kaggle 举办, 有 693 只队伍, 该竞赛内容是开发机器学习模型, 用于模拟气候模型中的次网格尺度大气物理过程, 以提升气候预测的精度并减少不确定性。
- 我们发现基线模型 1DUNet 精度较差, 为了提升性能, 我们做了以下尝试
 - 在网络结构上, 我们尝试了 Transformer、LSTM、GRU、1DCNN、UNet, 为 CNN 网络增加 SENet 模块, 并对上述模型进行组合
 - 在损失函数上, 我们尝试了 L1 和 L2 损失和带掩码的 L1、L2 损失
 - 在特征工程上, 我们对不同气压层的气象变量差分, 以及对不同气象变量组合相乘
- 我承担了 80% 的工作内容, 最终我们取得了 $R^2 = 0.76816$, 较基线模型提升了 17.1% 的精度, 取得了 35 名的成绩

2024 FLAME AI Challenge 竞赛, 个人

2024.9 - 2024.10

- 该比赛由 Kaggle 举办, 有 172 只队伍, 该竞赛内容是开发一个时空机器学习模型, 用于预测野火传播。基于高分辨率流场数据, 利用自回归方法学习野火传播的复杂动力学, 并预测未来火线位置。
- 在网络结构上, 我采用了 UNet, 并且尝试了修改了 UNet 的卷积模块 (如带残差连接的模块, 用 Vision Transformer 替换卷积模块和添加 Attention 模块)
- 在损失函数上, 我尝试了二元交叉熵损失和 MSELoss
- 在特征工程上, 我对不同时间步的变量差分, 以及对不同变量组合相乘
- 取得了 $MSE = 0.00987$, 较基线模型提升了 60.1%, 取得了第 4 名的成绩

专业技能

- 熟练使用 Python 进行开发编程, 熟悉 netcdf、tiff、npy、npz、csv、parquet 等气象数据格式的读写操作
- 熟悉 Linux 开发环境, 熟悉 Pytorch 深度学习框架, 熟悉 sklearn、lightgbm、xgboost 和 catboost 机器学习框架
- 熟悉 Transformer、LSTM、1DCNN 等时间序列深度学习算法
- 熟悉 UNet、ViT 等图像深度学习算法
- CET4 473 分, 雅思 5.5 分, 能够熟练阅读英文文献和技术文档