

基于Web的气象人工智能应用实习实训系统研发

■ 江元军 薛建军 丛藏程林

人工智能技术在气象领域的快速融合应用对高质量的气象教育培训提出了新的要求。以人工智能技术及应用培训需求为基础,提出了一种基于Web的气象人工智能实习实训系统并进行了设计与开发。本系统采用B/S架构,在Linux系统上利用人工智能领域Python生态系统的工具组件和核心库包构建,安装维护简单、高效集约、操作便捷,可面向多场景功能复用,适用于线上/线下技术培训,并在教学实践中得到了初步应用。

中图分类号: P4

文献标志码: A

DOI: 10.3969/j.issn.2095-1973.2025.01.009

当前,人工智能及其相关技术应用在气象领域迅速掀起了新一轮技术革新,2021年以来美国、欧盟、日本等国家或地区的一批重要气象机构纷纷出台战略规划、计划,不断强化人工智能技术在其业务和科学研究中的布局。中国气象局对此也做出了一系列部署和重点攻关方向上的工作安排。人工智能这一新技术、新方法正在快速融入气象业务和科研工作。

随着分层分类人工智能气象应用技术培训体系建设的不断深入,对培训班型开发和教学内容的精准化、精细化提出了新的要求。由于人工智能技术应用培训受到“算力”“数据”和“领域知识”等多重要素的驱动,要实现高质量的技术培训不仅需要完善的培训班型、精细的课程设计,而且还需要与之协调、匹配的实习实训环境做支撑和保障。本研究在跟踪调研国内外人工智能气象应用实习训练平台基础上,结合当前气象新技术培训实际情况,提出了一种基于Web的气象人工智能实习实训环境建设方案并完成了原型系统的研发。

1 现状分析

20世纪80年代以来,人工智能技术在气象领域进行了较多的探索性应用研究与实践,从气象资料的处理分析到预报产品的制作都有涉及。近年来,作为人工智能技术的一个分支,基于数据驱动的机器学习、深度学习等技术在气象多个领域取得突破性应用成果。随着人工智能通用算法、模型库的通用性、易用性不断提升,与领域知识密切相关的行业应用和技术培训受到了更多的关注。一些企业也纷纷提供或者构

建了面向人员技术培训的人工智能技术培训平台。比如,谷歌、微软、华为、百度等都提供基于云计算平台的机器学习、训练服务平台,有的还提供了少量人工智能训练数据集,可进一步实现简单的人工智能入门实操训练。但是这些解决方案和计算训练平台大多面向通用人工智能技术领域,多数系统基于大数据云平台建设,而对具有鲜明行业特色的气象领域关注不足,对气象领域中特定的需求适用性不高。因此,需要研发一套以满足当前和未来一段时间气象新技术培训为目标,安装维护简单、操作便捷、高效集约、多场景功能复用,同时适用于线上/线下技术培训的实习实训虚拟系统。

2 系统设计

2.1 总体架构

根据培训需求和技术路线调研成果,本系统采用B/S架构,以VMware为底层基础构件,CentOS操作系统为核心,集成Anaconda基础资源包,搭建以Python语言为基础生态,集成开发了包括常见气象资源库、人工智能机器学习环境。教师、学员可以通过Jupyter Notebook提供的统一用户交互界面,借助浏览器编程、调试、运行并显示结果,可以实现实习实训全过程在线操作并可根据实际需要可通过浏览器任意切换实训环境。图1给出了本系统的网络拓扑结构示意图。系统目前规划部署在气象网路DMZ区,优先使用本机算力和数据资源,同时预留接口通过气象业务内网获取相关算力和数据资源支持,实现算力、数据资源的

收稿日期: 2023年4月13日; 修回日期: 2023年8月4日

第一作者: 江元军(1968—), Email: q_jiang@sina.com

通信作者: 薛建军(1986—), Email: xuejianjun@cma.gov.cn

资助信息: 中国气象局气象干部培训学院科研项目(2022CMATCZD15); 中国气象局青年创新团队(CMA2023QN10); 中国气象局软科学重点项目(2023ZDIANXM08)

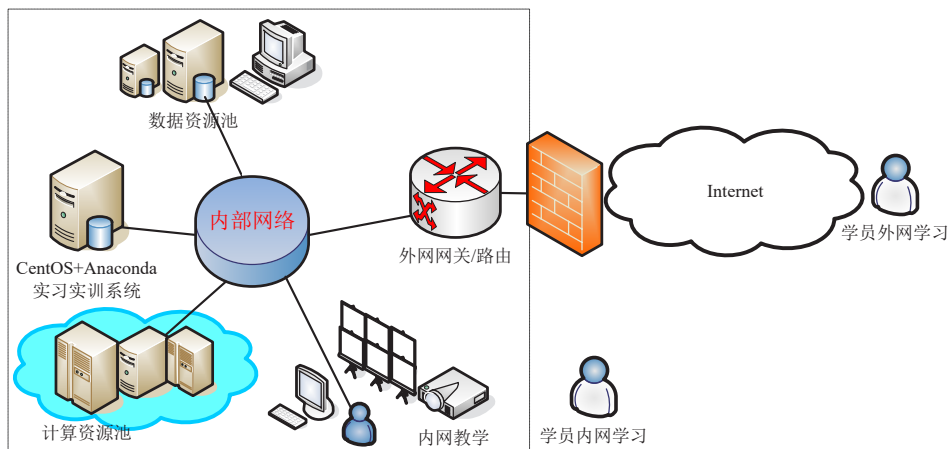


图1 系统的网络拓扑

扩展。系统通过Web访问方式，支持教师、学员内外网的在线实习实训和有关技术开发活动。

2.2 功能设计

系统按照功能结构可划分为基础资源设施、虚拟化平台、核心组件（库包）和教学应用四大部分（图2）。各部分主要功能如下：

1) 基础资源设施。构建系统所必需的数据、算力、网络通信和存储等数据和IT资源，该部分主要实现数据、算力等资源的统一接入和扩展。

2) 虚拟化平台。在基础资源设施上，采用VMware+CentOS+Anaconda的技术路线构建虚拟化平台，对下实现资源使用和调度，对上承载、部署相应的应用和服务等功能。

3) 核心组件（库包）。这是系统构建的核心内容。综合考虑现阶段气象新技术、新方法技术培训需

求，初步构建了以Python语言为主要基础环境，利用Jupyter Notebook作为统一集成开发工具和交互界面，实现对数据分析与可视化、气象应用、机器学习和数值模式等组件（库包）的搭建、管理、运行和调度等。具体内容见3.1节。

4) 教学应用。通过浏览器切换相应的实习实训环境，进行交互式在线学习训练，可以支撑机器学习、多源数据可视化、气象数据分析与预处理、数值模式及数值产品应用技术培训等多场景需求的气象新技术、新方法、新资料等实习实训。

3 系统实现

3.1 核心组件（库包）

当前在人工智能领域已经构建起了一套比较完善的基于Python生态系统的工具组件和相关库包。结合技术培训需求，通过对比分析多种技术方案，本系统

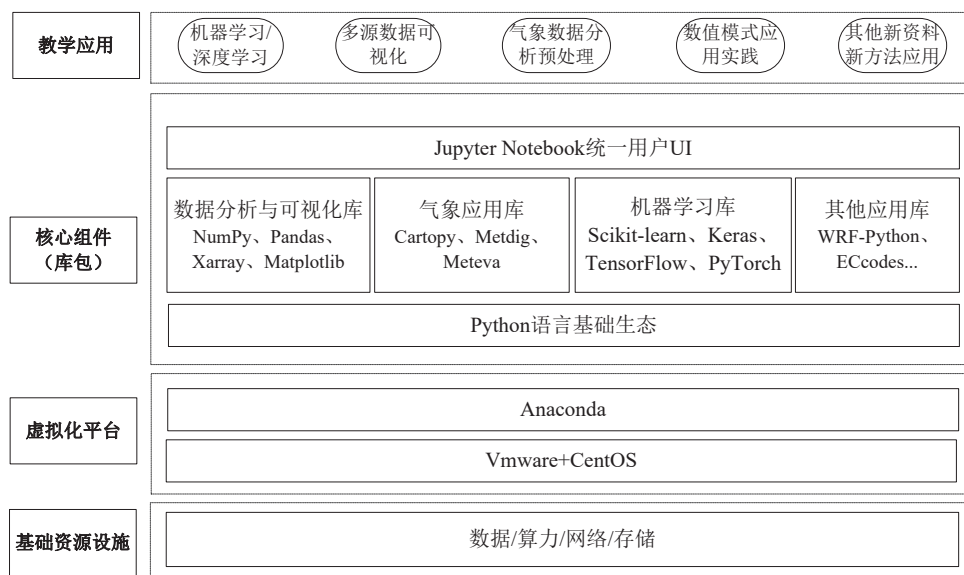


图2 系统功能结构

主要构建以下核心组件和库包:

3.1.1 Anaconda

由于Python具有功能完善的标准库和数量庞大的第三方库,应用领域范围遍及人工智能、数据分析、大数据及云计算等。通过对库的引用,能够实现对不同领域业务的开发。然而,正是由于库的数量庞大且各种库之间具有复杂的关联,对于管理这些库以及对库进行维护是一项既重要又复杂的工作,Anaconda是一个帮助管理各种库的平台,通过conda和pip管理各种库文件的安装、删除与升级,同时避免了各种库之间带来的冲突问题。

3.1.2 Jupyter Notebook

Jupyter Notebook是一个基于Web的开源交互式编程环境。支持运行包括Python在内的40多种编程语言,可以使用它来创建和共享包含实时代码、方程、可视化和文本的文档,是方便教学、自学Python的重要工具。

3.1.3 数据分析与可视化库

Python包含丰富、完善的第三方数据分析库,可以实现数据收集、预处理、可视化、数据模型训练、深度学习、自然语言处理等功能。NumPy、Pandas、Xarray和SciPy的协同工作可以高效解决很多问题,在天文学、生物学、气象学和气候科学,以及材料科学等多个学科得到了广泛应用。

1) NumPy是一个运行速度非常快的数学库,是Python科学计算的基础软件包,主要用于对数组执行各种数学运算,保证了数组和矩阵的高效计算,并内置了一个庞大的高级数学函数库,是构建Python科学计算大厦的基石。

2) Pandas是机器学习和数据分析中进行数据预处理的最重要的工具,它是基于NumPy开发的科学计算库,是一个强大的分析结构化数据的工具集,用于数据挖掘和数据分析,同时也提供数据清洗功能。

3) Xarray是处理多维数组的科学计算库,广泛应用于物理、天文、地球科学、生物信息、工程、金融、深度学习等领域,特别适合于处理气象中常用的netCDF文件。

4) Scipy是用于数学、科学、工程领域的常用软件包,可以处理最优化、线性代数、积分、插值、拟合、特殊函数、快速傅里叶变换、信号处理、图像处理、常微分方程求解等科学与工程中常用的计算。

5) Matplotlib提供了一套面向对象绘图的API,是Python中最受欢迎的数据可视化软件包之一,提供了常用的二维及简单的三维绘图接口,与NumPy、

Pandas一起使用,是数据分析和机器学习中不可或缺的重要工具之一。

3.1.4 常用气象应用库

1) Cartopy是专为地理空间数据处理而设计,以生成地图和其他地理空间数据分析, Cartopy包对Matplotlib包的功能进行了扩展,两者结合使用能绘制各种地图。

2) Metdig由国家气象中心天气预报技术研发室开发的面向国内天气预报业务和科研应用的通用型天气诊断分析工具包,为天气过程预报、复盘、机理研究等场景提供诊断分析技术支持。

3) Meteva是国家气象中心天气预报技术研发室开发的面向气象预报产品及中间产品进行全面检验评估需要用到的函数和功能的集成库,集合了数值模式和客观算法、精细化网格预报、各种应用预报产品等所需的各种检验算法。

3.1.5 机器学习库

1) Scikit-learn的功能主要分为六大部分:分类、回归、聚类、数据降维、模型选择和数据预处理。在气象上常用于图像识别、数据预处理等。

2) Keras是由谷歌工程师François Chollet开发的开源人工神经网络库,支持现代人工智能领域的主流算法,支持多操作系统下的多GPU并行计算,可以作为Tensorflow、Microsoft-CNTK和Theano的高阶应用程序接口,进行深度学习模型的设计、调试、评估、应用和可视化,被认为是深度学习开发端的主要工具。

3) TensorFlow是由Google Brain Team团队于2015年发布的完全基于Python用于大规模深度学习的开源库,被广泛运用在图像识别、图片分类、递归神经网络、自然语言处理、视频检测等深层神经网络等领域。支持CPU、GPU及多操作系统。

4) PyTorch是Meta(原Facebook)公司于2016基于Torch开发的Python开源库,是在支持GPU加速的张量(Tensor)计算和基于磁带(前向和反向)的梯度自动计算(Autograd)系统上搭建的深度学习神经网络,能够实现各种深度学习算法,如卷积神经网络、循环神经网络、变分自编码器等,同时具有强大的自动求导功能,是目前使用最广泛的机器学习框架之一,主要用于自然语言处理等,在气象领域亦被广泛应用。

3.1.6 其他常用气象库

包括netCDF4、metpy、metdig、cfrib、ECCodes、windrose、siphon、saalem、WRF-Python、geopandas、f90nm等是气象科学研究、数据处理和计算等常见库包。

3.2 系统搭建

在研究分析有关系统构建技术方案基础上, 基于简单、高效、易维护、成本低的原则, 本系统采用B/S架构, 在Linux服务器上安装Anaconda并配置人工智能虚拟环境。客户端利用本地Web浏览器通过url访问远程服务器上Anaconda的Jupyter Notebook服务, 实现了在本地浏览器上对远程服务器上的Python程序进行编辑、运行以及任意切换并管理各种虚拟环境的功能。学员无需配置环境便可直接动手练习, 大幅度减少了配置环境所需的学习成本和时间成本。在进行实习实训时可通过以下操作完成既定任务。

1) 文件管理。通过Web浏览器可以轻松管理文件(夹), 实现新文件(夹)的创建、编辑与运行, 以及文件的上传与下载。

2) 实时显示当前程序运行情况, 并可随时中断程序的运行。

3) 环境管理。可以在Web浏览器中安装、删除、更新各种第三方模块(库)、虚拟环境, 并可以自由切换各种实习环境。

此外, 在教师管理端还提供服务器集群管理功能, 显示服务器集群的运行状况, 根据实际需要可以随时启动或关闭服务器集群节点, 调整算力等资源支持。

3.3 实现功能

基于最大限度利用资源的原则, 本平台采用VMware+CentOS+Anaconda模式, 在原型系统中实现了以下功能。

3.3.1 创建配置虚拟环境

平台创建了Python基础、气象诊断分析、Pytorch和TensorFlow-keras四个虚拟环境, 并可根据教学需要随时增加相应的教学虚拟环境, 其中气象诊断分析环境可以实现常规气象数据(站点、雷达、卫星、数值预报产品)等的分析、绘图等功能, 需要安装气象相关的软件库netCDF4、cfrib、cartopy、metpy、ECcodes、WRF-Python、meteva、metdig等, 需要指出的是在配置气象诊断分析环境时需使用Python3.8或3.7版本。由于Keras是基于TensorFlow的深度学习库, 可以看作是TensorFlow封装后的一个API, 因此将这两个机器学习库安装在一个环境内, 即TensorFlow-Keras。

3.3.2 实现利用Web自由切换虚拟环境

在虚拟环境中使用Jupyter Notebook编辑运行程序, 一种方案是先在后台进入虚拟环境, 然后在虚拟环境中运行Jupyter Notebook。考虑到这种方法比较繁

琐, 在每一个虚拟环境中安装ipykernel库并在base环境中配置了nb_conda库, 从而实现了在Web上自由切换虚拟环境的功能, 使系统操作更简洁, 用户体验更便捷(图3)。

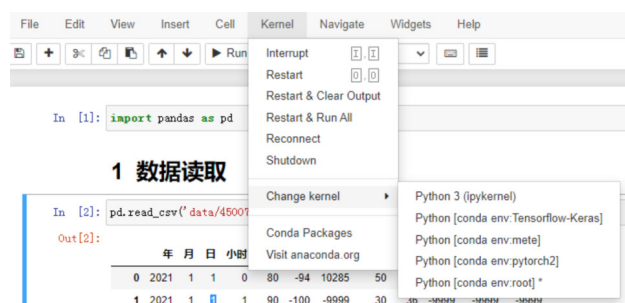


图3 虚拟环境切换

3.3.3 系统性能测试

本系统的性能主要依赖于硬件条件。目前系统安装在一台ThinkSystem SR850 (64 CPUs×Intel(R) Xeon(R) Gold 6130, 内存 223.69 GB) 服务器, 测试显示, 在Python及其相关的气象应用培训中, 可以满足10人左右的在线应用需求, 对于需要更多算力和存储支持的机器学习训练可以支持1~2个基础模型训练。但是凭借系统开放的设计架构, 随着硬件的扩充升级, 特别是借助服务器集群超强的算力和存储能力, 系统性能将得以显著提升。

4 应用实例

在Jupyter Notebook统一用户界面环境下, 教师利用Web浏览器对程序代码逐行编写、展示讲解的同时进行代码运行, 并显示运行结果, 系统实现了人工智能教学实例从数据读取、模型构建、选择损失函数和优化器、模型训练到图形化展示的实习实训全流程, 与此同时, 学员可以在自己的用户端上跟着教师一起同步操作, 实现教师教学演示与学员实践操作较好地配合, 显著提升了教学和学习体验, 在实际培训应用中得到了教师和学员充分肯定。

以下是利用Pytorch深度学习框架, 对地基云雷达、FY-4A卫星数据进行训练, 生成高精度云顶高产品的一个应用实例。该实例针对地基云雷达观测数据精度高, 但点状、稀疏而卫星数据空间覆盖范围大, 但云顶高精度相对较低的特点。首先对地基云雷达回波分析获得准确的云内宏观参数, 如云厚、云高、云量等, 通过反演获得云内微物理参量, 如云滴半径、云滴数浓度、滴谱、冰与液态水的含量等; 其次将FY-4A卫星数据和地基雷达两种数据进行时空匹配, 建立数据集; 最后, 经过模型训练, 生成空间覆盖范围广、精度相对高的云顶高产品(图4)。

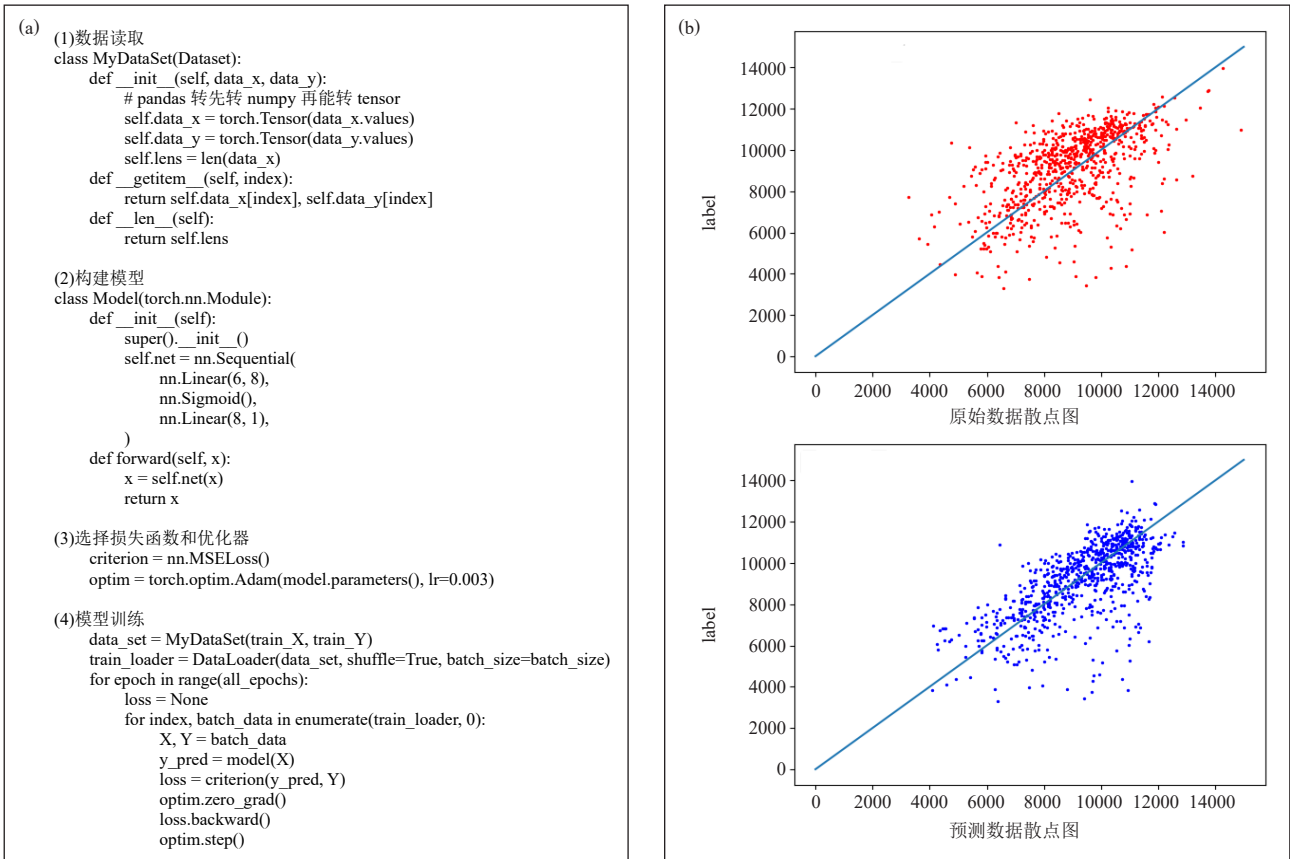


图4 Pytorch模型训练代码示例(a)以及训练结果(b)

5 结语

以人工智能技术培训需求为基础，提出了基于Web的气象人工智能实习实训环境建设方案并进行了原型系统研发，构建了一个操作简便、易于扩展维护、可面向多场景功能复用统一的人工智能实习实

训环境，减少教学实践中复杂的环境搭建，帮助教师、学员快速进入教学训练活动。下一步，还将随着相关技术的发展以及教学实践的反馈继续优化和完善系统。

深入阅读

陈华, 2022. 基于Python的数据分析与可视化平台研究[J]. 网络安全技术与应用(2): 57-58.

杜智涛, 姜明波, 杜晓勇, 等, 2021. 机器学习在气象领域的应用现状与展望[J]. 气象科技, 49(6): 930-941.

段炼, 岳炼, 张展硕, 等, 2020. 基于Python的WRF模式后处理研究[J]. 信息技术与信息化(9): 50-53.

国威, 2020. 人工智能背景下的Python教学策略研究[J]. 电脑知识与技术, 16(31): 174-176.

国务院, 2022. 气象高质量发展纲要(2022—2035年)[EB/OL]. (2022-04-28)[2023-04-10]. https://www.gov.cn/zhengce/zhengceku/2022-05/19/content_5691116.htm.

胡永波, 2021. 人工智能云服务模式的系统架构分析[J]. 集成电路应用, 38(5): 98-99.

气象史料挖掘与研究工程项目组, 2018. 20世纪80年代人工智能在天气预报中的应用[J]. 气象科技进展, 8(3): 88-91.

王亚东, 2017. Python在气象数据可视化中的应用[C]//第34届中国气象学会年会S20气象数据: 深度应用和标准化论文集. 郑州:

中国气象学会, 139-143.

吴灿, 戴洋, 何晓欢, 等, 2021. 气象和大气科学领域人工智能科学研究的国际态势分析[J]. 科学咨询(46): 1-5.

熊华东, 莫华, 李登峻, 2009. 基于Web环境下教学资源平台的研发[J]. 江西电力职业技术学院学报, 22(4): 65-66, 72.

许小峰, 2018. 从物理模型到智能分析——降低天气预报不确定性的新探索[J]. 气象, 44(3): 341-350.

薛建军, 贾朋群, 肖子牛, 2022. 集合、同化思想在大气科学中的渗透[J]. 气象科技进展, 12(6): 64-72.

张萌, 贾朋群, 王小光, 2020. 人工智能技术在大气科学领域的应用及其发展态势[J]. 科学观察, 15(1): 22-29.

中国气象局, 2021. 加强气象科技创新工作方案[EB/OL]. (2021-04-20)[2023-04-10]. https://www.cma.gov.cn/2011xwzx/2011xqxxw/2011xqxxyw/202104/t20210420_575524.html.

中国气象局, 国家发展改革委, 2021. 全国气象发展“十四五”规划[EB/OL]. (2021-12-02)[2023-04-10]. https://www.cma.gov.cn/2011xwzx/2011xqxxw/2011xqxxyw/202112/t20211224_4325497.html.

(下转78页)

深入阅读

- 吴楠, 2024. 南京大学发布人工智能通识核心课方案[EB/OL]. (2024-03-07)[2024-04-01]. <https://www.nju.edu.cn/info/1056/357721.htm>.
- 许小峰, 2018. 从物理模型到智能分析——降低天气预报不确定性的新探索[J]. 气象, 44(3): 341-350.
- 中国气象局, 2023. 人工智能气象应用工作方案(2023—2030年)[EB/OL]. (2023-07-28)[2024-01-02]. https://www.cma.gov.cn/2011xwzx/2011xqxxw/2011xqxyw/202307/t20230728_5677975.html.
- 中华人民共和国教育部, 2019. 教育部关于公布2018年度普通高等学校本科专业备案和审批结果的通知[R]. 北京: 教育部.
- Bi K F, Xie L X, Zhang H H, et al, 2023. Accurate medium-range global weather forecasting with 3D neural networks[J]. Nature, 619(7970): 533-538.
- Bouallège Z B, Clare M C A, Magnusson L, et al, 2024. The rise of data-driven weather forecasting: A first statistical assessment of machine learning-based weather forecasts in an operational-like context[J]. Bulletin of the American Meteorological Society, 105(6): E864-E883.
- Boukabara S A, Krasnopolsky V, Penny S G, et al, 2021. Outlook for exploiting artificial intelligence in the earth and environmental sciences[J]. Bulletin of the American Meteorological Society, 102(5): E1016-E1032.
- Charlton-Perez A J, Dacre H F, Driscoll S, et al, 2024. Do AI models produce better weather forecasts than physics-based models? A quantitative evaluation case study of Storm Ciarán[J]. npj Climate and Atmospheric Science, 7(1): 93.
- Chen L, Zhong X H, Zhang F, et al, 2023. FuXi: A cascade machine learning forecasting system for 15-day global weather forecast. npj Climate and Atmospheric Science, 6: 190.
- ECMWF, 2024. Training course: Machine learning for weather prediction[EB/OL]. (2024-03-02)[2024-04-01]. <https://events.ecmwf.int/event/377/>.
- ECMWF, 2023. MOOC machine learning in weather & climate[EB/OL]. (2023-09-01)[2024-01-02]. <https://lms.ecmwf.int/pages/index.html>.
- ECMWF, 2021. Machine learning at ECMWF: A roadmap for the next 10 years[R]. Shinfield Park: European Centre for Medium-Range Weather Forecasts.
- EUMETSAT, 2021. Artificial intelligence (AI) for earth monitoring[EB/OL]. (2021-10-22)[2024-01-02]. <https://www.futurelearn.com/courses/artificial-intelligence-for-earth-monitoring>.
- Irrgang C, Boers N, Sonnewald M, et al, 2021. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science[J]. Nature Machine Intelligence, 3(8): 667-674.
- NOAA, 2019. 1st workshop on leveraging AI in the exploitation of satellite earth observations & numerical weather prediction[EB/OL]. (2019-04-23)[2024-01-02]. https://www.star.nesdis.noaa.gov/star/meeting_2019AIWorkshop.php.
- NOAA, 2022. 4th NOAA workshop on leveraging artificial intelligence in environmental sciences[EB/OL]. (2022-09-06)[2024-01-02]. <https://www.noaa.gov/ai/events/4th-noaa-ai-workshop>.
- NOAA, 2021. NOAA artificial Intelligence Strategic Plan 2021—2025. [EB/OL]. (2021-01-01)[2024-01-02]. <https://repository.library.noaa.gov/view/noaa/56881>.
- Reichstein M, Camps-Valls G, Stevens B, et al, 2019. Deep learning and process understanding for data-driven Earth system science[J]. Nature, 566(7743): 195-204.
- Schneider T, Behera S, Boccaletti G, et al, 2023. Harnessing AI and computing to advance climate modelling and prediction[J]. Nature Climate Change, 13(9): 887-889.
- Sun Z H, Sandoval L, Crystal-Ornelas R, et al, 2022. A review of earth artificial intelligence[J]. Computers & Geosciences, 159: 105034.
- WMO, 2021. Future of weather and climate forecasting[R]. Geneva: WMO.

(作者单位: 中国气象局气象干部培训学院)
(编辑: 卢冰)

(上接74页)

深入阅读

- 中国气象局, 科学技术部, 中国科学院, 2022. 中国气象科技发展规划(2021—2035年).[EB/OL]. (2022-02-28)[2023-04-10]. http://sc.cma.gov.cn/zfxxgk/fdzdgnr/ghjh_85551/202203/t20220303_4555674.html.
- 中国气象局科技与气候变化司, 2022. 业务能力提升关键科技问题清单[EB/OL]. (2022-03-28)[2023-04-10]. https://www.cma.gov.cn/zfxxgk/gknr/wjgk/qtwj/202203/t20220328_4671396.html.
- 周康辉, 郑永光, 韩雷, 等, 2021. 机器学习在强对流监测预报中的应用进展[J]. 气象, 47(3): 274-289.
- Boukabara S A, Krasnopolsky V, Penny S G, et al, 2021. Outlook for exploiting artificial intelligence in the earth and environmental sciences[J]. Bulletin of the American Meteorological Society, 102(5): E1016-E1032.
- Irrgang C, Boers N, Sonnewald M, et al, 2021. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science[J]. Nature Machine Intelligence, 3(8): 667-674.
- JMA, 2018. JMA's NWP strategic plan toward 2030[EB/OL]. [2023-04-10]. https://www.jma.go.jp/jma/en/Publications/JMA_NWP_Strategic_Plan_Toward_2030.pdf.
- NOAA, 2021. NOAA Artificial Intelligence Strategic Plan 2021—2015[EB/OL]. [2023-04-10]. <https://repository.library.noaa.gov/view/noaa/56881>.
- Reichstein M, Camps-Valls G, Stevens B, et al, 2019. Deep learning and process understanding for data-driven Earth system science[J]. Nature, 566(7743): 195-204.
- Xiong A Y, Liu N, Liu Y J, et al, 2022. QpefBD: a benchmark dataset applied to machine learning for minute-scale quantitative precipitation estimation and forecasting[J]. Journal of Meteorological Research, 36(1): 93-106.

(作者单位: 中国气象局气象干部培训学院)
(编辑: 卢冰)