

时空大数据与人工智能助力地学领域复杂系统研究

于龙龙, 李根

成都大学 计算机学院, 四川 成都 610106

DOI:10.61369/ME.2025050035

摘 要 : 人类正面临全球气候变化与资源能源紧缩等“地学复杂系统”的叠加挑战。时空大数据在多源要素、跨尺度耦合与过程表征方面具有天然优势,但其价值释放仍受数据孤岛、质量不稳和算力门槛所限。本文系统梳理了时空大数据与人工智能在地学复杂系统研究中的应用现状、关键挑战与发展趋势,重点阐释从时空数据驱动到智能分析、再到平台化整合的技术迭代逻辑。针对数据瓶颈、协同门槛与工程实施等问题,提出以高性能计算、云计算与时空大数据平台为支撑的解决思路,并展望以时空图神经网络为代表的前沿技术融合方向。研究旨在为依托时空大数据的地学复杂系统研究提供可操作的理论框架与实践路径,推动其战略价值的充分释放。

关 键 词 : 时空大数据; 高性能计算技术; 机器学习; 时空图神经网络

Advancing Geoscience Complex-Systems Research with Spatiotemporal Big Data and AI

Yu Longlong, Li Gen

College of Computer Science, Chengdu University, Chengdu, Sichuan 610106

Abstract : Humanity faces compounding challenges in “geo-complex systems,” including global climate change and tightening resource and energy constraints. Spatiotemporal big data offers inherent advantages for integrating multi-source factors, linking processes across scales, and characterizing system dynamics, yet its potential remains constrained by data silos, quality instability, and computational barriers. This paper provides a systematic review of the current applications, key challenges, and emerging trends of spatiotemporal big data and artificial intelligence in geo-complex systems. We emphasize the technological progression from data-driven spatiotemporal analysis to intelligent modeling and, ultimately, platform-level integration. To address bottlenecks in data quality, cross-domain collaboration, and engineering deployment, we propose solution pathways grounded in high-performance computing, cloud computing, and spatiotemporal data platforms, and we outline convergence directions involving frontier methods such as spatiotemporal graph neural networks. The study offers an operational theoretical framework and practical roadmap for research anchored in spatiotemporal big data, facilitating fuller realization of its strategic value.

Keywords : spatiotemporal big data; high-performance computing; machine learning; spatiotemporal graph neural networks

引言

数据的根本使命在于消除不确定性。相较传统数据,时空大数据兼具时间与空间属性,在刻画地学系统的多尺度耦合过程方面更具优势,可以更为显著地消除地学复杂系统的不确定性。时空大数据、人工智能与物联网等技术的协同爆发为人类解决地学复杂系统中与自身命运紧密相关的重大问题提供了前所未有的基于。然而,时空大数据在地学复杂系统的应用当前仍面临数据、算力与技术的多维制约和关键挑战。本文系统探讨了时空大数据与人工智能在地学复杂系统研究中的应用现状、关键挑战和发展趋势,重点分析了从时空大数据驱动到智能分析再到时空大数据平台整合的技术迭代逻辑。同时,针对数据瓶颈、协同门槛以及技术实施等挑战,提出高性能计算、云计算和时空大数据平台等技术解决方案,并展望了时空图神经网络等前沿技术的融合趋势。

作者简介:

于龙龙,男,汉族,四川成都人,博士,成都大学计算机学院,数据科学与大数据系讲师。研究方向:时空大数据分析;

李根,男,汉族,四川成都人,博士,成都大学计算机学院,数据科学与大数据系讲师。研究方向:时间序列分析与人工智能应用。

一、时空大数据发展的新趋势

（一）时空大数据的巨大潜能

1. 时空大数据的爆发

广义上讲，具有时间和空间属性的数据就属于时空大数据，这一类数据体量极其巨大。从狭义上讲，时空大数据是带有时间信息的地理数据，包括长达40余年的遥感观测数据、地理空间数据集、生态监测数据等。随着带有时空属性的监测和数据记录平台、终端设备、传感器和物联网设施的爆发式增长，时空大数据会以越来越惊人的速度持续堆积。目前来看，时空大数据的潜能远远未能充分释放，特别是在地学极复杂系统的研究方面，例如全球碳循环响应机理结构、极端气象事件预测以及地球资源智能管控方面，这些已积累的时空大数据是解决这些领域关键挑战的战略性资源^[1]。

2. 既有模式的挖掘

通过学习已有的时空大数据中蕴含的特定模式和规律，可以为我们了解地学系统的复杂机理和演变规律提供指导，从而在关系人类命运的一系列地学复杂系统关键科学问题上提供重要的借鉴和参考，这种参考意义不仅包括对历史数据的研究得出模式、对现有数据分析得出现状，更包括对未来趋势的预判、事件的预测以及现实决策的指导。例如，研究人员通过研究新型的叶绿素荧光信号数据，了解全球植被生产力动态，进一步研究植被—气候响应机制，从而助力全球碳循环机理的研究以及双碳目标的实现^[2]。

3. 与大数据前沿技术的整合

时空大数据具有体量大、模态多、更新快与空间结构复杂等特征，超出了传统集中式与批处理范式的承载能力。近年来，分布式与并行计算、云原生与容器编排、以及流批一体的计算框架，为高效处理提供了可行路径：在存储层，数据湖/湖仓一体与列式存储结合时空索引（如空间填充曲线与瓦片化策略），显著缩短检索与扫描路径；在计算层，Flink/Spark 等支持增量计算与窗口化聚合，可在事件发生时保持模型与特征的在线更新，并通过 GPU/TPU 加速缩短训练—推理闭环；在治理层，元数据目录、语义本体与数据血缘使跨源数据“可发现、可解释、可追溯”，为跨任务复用与结果审计奠定基础。与此同时，云计算的弹性供给与边缘侧“近源计算”可按需调度算力，降低回传与等待成本，配合无服务器（Serverless）与按量计费机制，在时效与预算之间取得可验证的平衡。进一步地，特征库与模型注册中心强化了特征—模型—服务的一致性，支撑从离线验证到在线发布的 MLOps 流程；而联邦学习与差分隐私等合规技术，为跨机构、跨区域的数据协作提供了可执行的安全边界。综合来看，这些前沿技术的协同应用，使“数据—算法—平台”形成可迭代的闭环，从而将时空大数据的“可用性”转化为面向复杂地学问题的“可用能力”。

（二）人工智能和图神经网络的应用

人工智能助力时空大数据分析，并在数据质量改善、深层特征提取、复杂模式捕获等方面表现突出。在地学极复杂系统的研

究方面，二者的联合使得一些在单一学科和传统数据分析模式中很难解决的问题迎来空前的机会。如重大气象灾害预测，碳循环响应机制分析等。其中，图神经网络是一种能够处理图结构和时空大数据的深度学习模型，近年来已成为新型非显式特征学习的研究热点。与传统神经网络主要处理规则化数据不同，图神经网络擅长处理具有复杂拓扑结构的非欧几何数据，例如社交网络、知识图谱、生物分子结构和交通网络等。图结构中的节点表示实体，边表示实体间的关系，图神经网络通过学习节点与其邻居节点之间的相互作用，充分利用图的结构信息，从而实现对复杂数据的有效建模。早期的图神经网络主要包括图卷积网络和图注意力网络等。在时空大数据领域，图神经网络被用来分析遥感图像中地物的空间关系。随着时空数据的重要性不断提升，研究者们开始将图神经网络应用于时空数据分析中，发展出时空图神经网络等前沿技术，进一步推动了时空大数据在地学复杂系统研究中的应用。

二、时空大数据潜能充分释放的挑战

时空大数据在破解地学复杂系统的关键重大问题上的价值日益凸显，但其价值开发和潜力释放的速速明显不及其数据量的爆发速度。重要原因在于如下几方面的挑战。

（一）数据质量瓶颈

时间分辨率、空间分辨率、时间跨度与空间覆盖的不匹配，直接影响单源数据的应用可行性与多源数据的时空一致性；云雨遮挡、轨道漂移与器件老化会在数据生成端引入系统误差；地表异质性与观测噪声进一步放大不确定性；人类活动数据还伴随隐私约束与抽样偏倚。若不在数据源与治理环节及时处置，后续建模将不可避免地积累偏差并传导至决策层。可行对策包括：跨传感器交叉定标与偏差订正、在物理/统计一致性约束下的时空插补与重建、质量标注与置信度评估，以及对关键要素实施持续的漂移监测与再采样校正。此外，针对现有数据集，依托人工智能和高性能计算进行数据质量优化也是一大迫切需求。

（二）数据协同门槛

不同数据在组织模式与格式上的差异显著，且常伴随坐标基准/投影不一致、时间基准与采样策略不统一、接口协议异构，以及数据许可与访问控制体系各自为政等问题。这些因素叠加，造成语义断裂与流程碎片化，使跨域协同成本长期居高、经验难以迁移。为此，在技术层面需要建立统一的要素本体与中间语义层，完善元数据编目、特征/标签接口、数据血缘与版本化管理，并实现跨库可发现与权限协同；在制度层面需要明确共享协议、合规边界与评价激励，形成稳定的协作机制。更为关键的是，面向地学复杂系统的综合问题，研究者需具备问题拆解与跨语域映射能力，并能在学科模型与工程实现之间完成有效整合。

（三）时空大数据高效驱动的技术和成本问题

时空大数据面向采集、预处理、统计学习、智能建模与结果分析的全生命周期提出海量且连续的计算需求，传统集中式方案难以同时满足低延迟与高吞吐。实践中，“时效—稳定—成本”

的权衡成为首要约束：一方面，在线更新、漂移监测与多源对齐要求弹性算力与可观测机制；另一方面，长期存储与频繁重算又显著抬高预算。此外，跨学科地学模型与 AI 算法的工程化落地存在较高门槛，实时性要求（例如分钟级预警）进一步放大了调度、容错与一致性控制的复杂度。

三、未来发展趋势与战略建议

（一）时空图神经网络

时空数据具有复杂的空间和时间依赖性，例如气象数据、交通流量、遥感观测等，传统机器学习方法难以充分捕获这种复杂的时空关联性。随着时空数据资源战略价值的不断提升与日益凸显，研究者们发展出时空图神经网络（Spatiotemporal Graph Neural Networks, ST-GNN）。ST-GNN 通过在图神经网络中引入时序建模机制，如循环神经网络、长短时记忆网络以及时序卷积网络，可以更好地捕获数据的时空动态特征。ST-GNN 模型利用注意力机制更好地捕获时空数据中的关键特征。这些模型在交通预测、环境监测、空气质量预报等领域取得了优异的效果。ST-GNN 在环境、生态系统监测等领域的成功应用，表明其在多源数据融合、时空数据建模和数据重建等方面具备巨大潜力。时空图神经网络在未来时空大数据分析，特别是在地学复杂系统关键重大问题的解决中，必然发挥越来越重要的作用^[3]。

（二）高性能计算技术的快速发展

并行计算、分布式计算、云计算平台和超级计算资源的使用门槛降低，越来越普及。Google colab、阿里云、国家超算中心等，为海量时空大数据的高效处理、人工智能技术的应用提供了计算实现框架。极端事件对延迟与吞吐提出并行要求。可行策略包括：构建弹性算力池，将按需/抢占式实例与本地/云端算力统一编排；实施成本感知调度，以紧迫度、SLA 与预算为多目标优化约束；推进数据近源计算与流批一体，在边缘侧完成预处理与快速筛选，核心计算在云/超算侧完成；建立可观测性体系，将延迟、吞吐、错误率与模型漂移纳入统一监控，并支持灰度发布与滚动升级^[4]。

（三）时空大数据平台

将时空大数据资源、高性能计算框架、云计算环境和先进的

人工智能技术进行平台化整合，不断降低时空大数据智能分析的门槛，已成为一大必然趋势。以 GEE 为代表的云计算平台，已经显示出了突破传统时空大数据分析模式和技术瓶颈的巨大优势，极大促进了对地学复杂系统关键问题的研究。LuoJia-net 等新兴时空大数据平台进一步整合了地学算子与模型库，能更好地服务于地学复杂问题的解析。在时空大数据平台的工程化落地方面，如下的问题需进一步优化。一是接口标准化（OGC/WMS/WFS），以解决接入与编排的一致性。二是生态分层（数据—算子—模型—应用），以支撑能力复用与组合创新；三是可追溯治理（血缘与版本化管理），以确保结果审计与合规。在多租户、多任务场景下，还需提供更加智能的资源隔离、配额与成本感知能力。时空大数据平台不仅需要引入人工智能模型接口和应用场景，其本身的构建和运行也可以借助人工智能模型来优化^[5]。

（四）更为紧密的学科交叉

地学复杂系统本身具有高度的学科交叉性，单一视角和单一领域的的数据难以有效结构其中的复杂机理，因此学科的进一步交叉融合是必然趋势。通过各学科领域数据集的协同以及前沿的人工智能方法如时空图神经网络的应用，以及高性能计算和云计算平台的赋能，时空大数据这一战略资源才能更好地为人类解决地学复杂系统中与自身命运密切相关的重大问题服务。

四、结论与展望

时空大数据的爆发式增长与人工智能的快速演进，为地学复杂系统研究提供了前所未有的机遇。伴随高性能计算、云计算与时空大数据平台的迭代升级，数据获取、治理与分析的时间与工程门槛持续下降，方法体系日趋成熟并可复制到更多场景。尽管数据质量、跨域协同与实时部署仍构成关键挑战，三者与算力调度、隐私合规的协同优化，有望在“时效—稳定—成本”之间形成可验证的平衡。面向未来，应推进机理—数据融合的算法体系、建立跨区域可复现开放基准，完善从数据到服务的可追溯工具链与人才/组织协作机制。可以预期，时空大数据、人工智能与高性能计算的深度整合，将显著提升人类在气候变化、极端灾害与资源紧缩等重大议题上的主动性与应对能力。

参考文献

- [1] 王家耀, 王利军, 程士源. 时空大数据及其应用 [J]. 测绘科学技术学报, 2024, 40(04): 331-337+391.
- [2] Yu L, Wen J, Chang C Y, et al. High-resolution global contiguous SIF of OCO-2[J]. Geophysical Research Letters, 2019, 46(3): 1449-1458.
- [3] Wu Z, Pan S, Chen F, et al. A comprehensive survey on graph neural networks[J]. IEEE transactions on neural networks and learning systems, 2020, 32(1): 4-24.
- [4] 李宇航, 徐志伟, 刘燕华, 等. 人工智能时代的地理科学前沿问题探析 [J]. 地理学报, 2024, 79(10): 2409-2424. DOI:CNKI:SUN:DLXB.0.2024-10-001.
- [5] Tamiminia H, Salehi B, Mahdianpari M, et al. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review[J]. ISPRS journal of photogrammetry and remote sensing, 2020, 164: 152-170.