

每日小结

	周一	周二	周三	周四	周五
早	Scala、spark 集群语言处理访问数据	阅读文献，写周报	组会	访问数据筛选、统计，阅读文献	Scala、spark 集群语言处理访问数据
中	访问数据筛选、统计	Scala、spark 集群语言处理访问数据	阅读文献	阅读文献，graph 建立代码	访问数据处理后整理，阅读文献
晚	阅读文献，GNN 代码	访问数据筛选、统计，写周报		Scala、spark 集群语言处理访问数据	访问数据筛选、统计

注：简单表述当前时间段工作，如看文献 1，整理数据等

科研详情

文献阅读

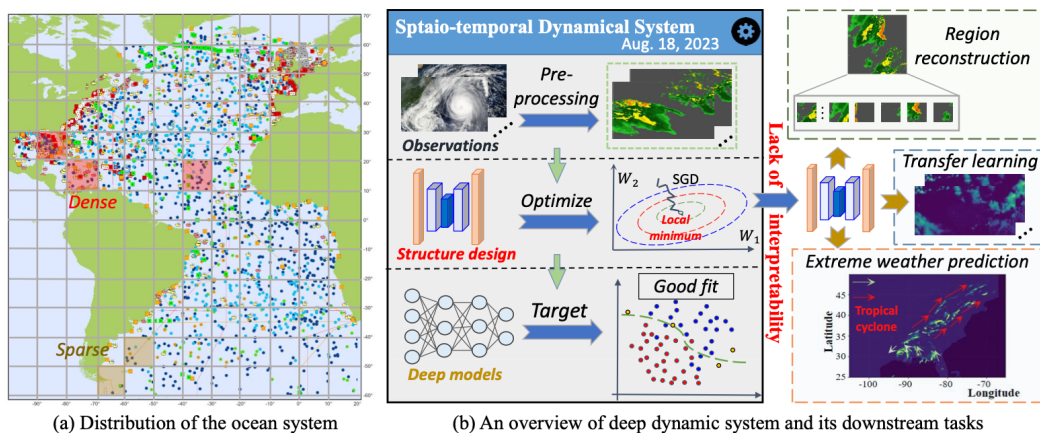
文献 1

题目：NuwaDynamics: Discovering and Updating in Causal Spatio-Temporal Modeling

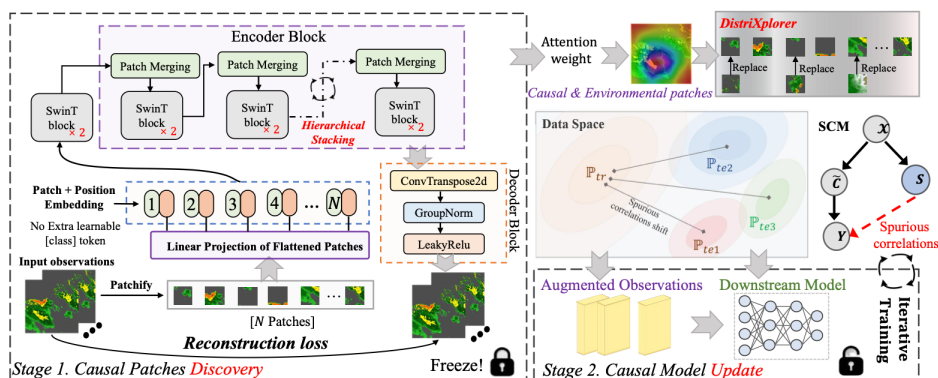
作者：Kun Wang, Hao Wu, Yifan Duan, Guibin Zhang, Kai Wang, Xiaojiang Peng, Yu Zheng, Yuxuan Liang, Yang Wang

出处：ICLR 2024

方法：



提出了一种因果时空建模 NuwaDynamics，用于识别数据中的因果关系并为模型注入因果推理能力。通过自我监督和干预，在上游发现阶段注入广义信息，然后将数据传递到下游任务以帮助模型识别更广泛的潜在分布和培养因果感知能力。实验证明，集成了 NuwaDynamics 概念的模型在不同任务上取得了显著成果，比如极端天气和长时间步长超分辨率预测。



NuwaDynamics 包含 Discovery 和 update 两个阶段。使用 Swin Transformer 作为上游模型。NuwaDynamics 的因果：根据因果理论，通过四个变量构建了结构因果模型（SCM）（Pearl, 2009）：输入观测值 X 、真实值 Y 、 X 中表示为 \tilde{C} 的因果块以及表示的混杂因素（即环境）。通过以下方式来描述它们之间的因果关系：

- $\tilde{C} \leftarrow \mathcal{X} \rightarrow S$. The input \mathcal{X} consists of two disjoint parts \tilde{C} and S .
- $\tilde{C} \rightarrow \mathcal{Y} \leftarrow S$. \tilde{C} is the only endogenous parent to determine the ground-truth \mathcal{Y} . However, in practical scenarios, S is simultaneously used for predicting \mathcal{Y} , leading to spurious associations.

下游模型可以分为 Transformer 类和非 Transformer 类。对于 Transformer 类来说，保证上下游模型的一致性可以实现参数的快速传递，有利于下游模型更快的优化。另一方面，对于 Transformer 架构，采用传输增强数据来优化和更新下游模型，提高其因果感知能力。

文献 2

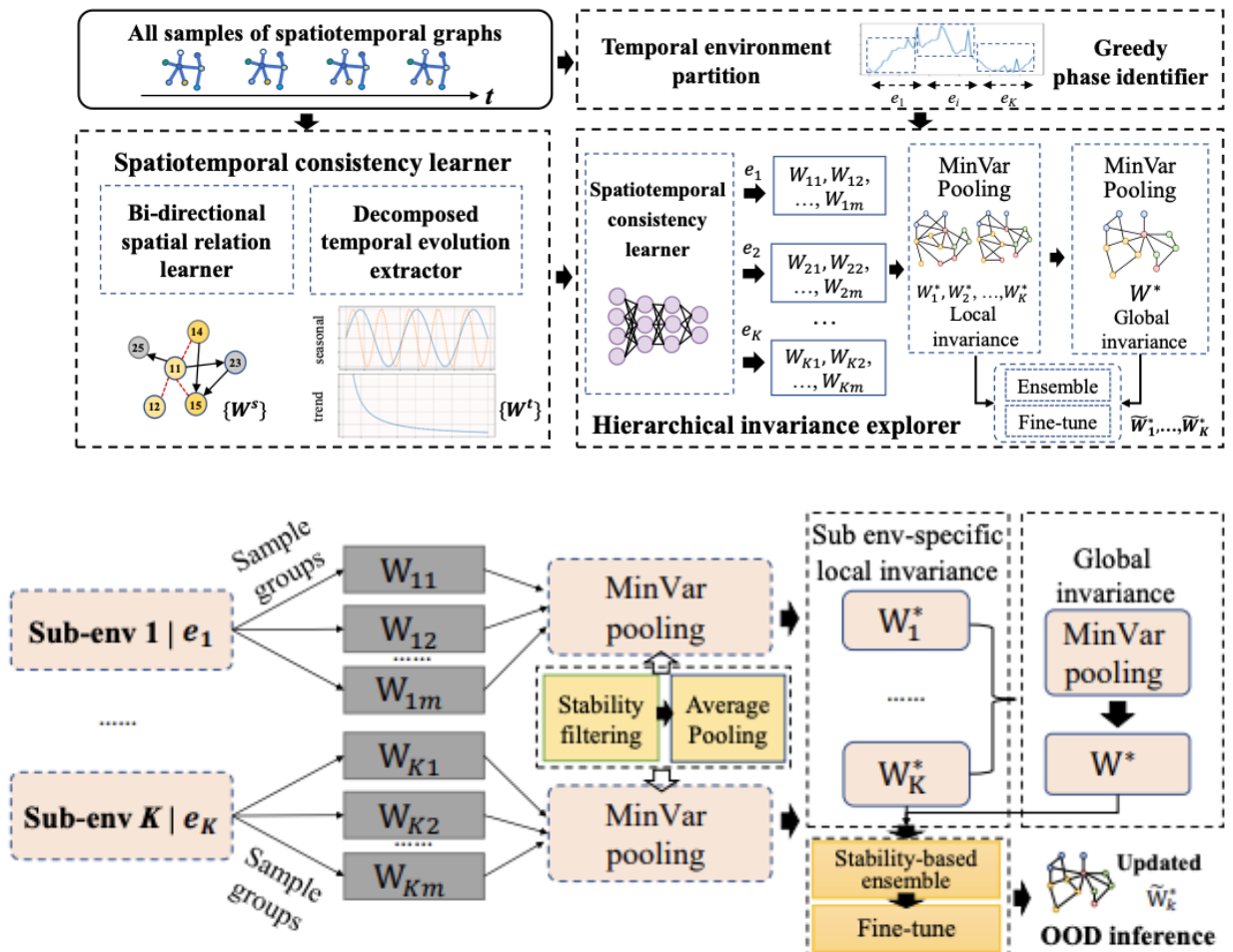
题目：Maintaining the Status Quo: Capturing Invariant Relations for OOD Spatiotemporal Learning

作者：Zhengyang Zhou, Qihe Huang, Kuo Yang, Kun Wang, Xu Wang, Yudong Zhang, Yuxuan Liang, Yang Wang

出处：KDD '23: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2023

方法：

本文提出了因果时空学习框架 CauSTG。它通过将时间步骤视为环境，并将时空关系转换为学习参数，使不变关系适用于 OOD 场景。本文将时间步骤作为环境，并将时空关系转化为可学习的参数。为了解决周期性的异质性，本文通过识别不同的趋势模式将时间步骤划分为子环境，从而使重新组织的样本能够单独训练。CauSTG 量化了空间一致性，提取了季节性趋势模式，并使用基于变异的过滤器实现局部和全局不变性。



为了提取 ST 观测中的不变性，作者提出了一个时空一致性学习器和一个分层不变性学习器来联合过滤稳定关系。时空学习器量化双向空间一致性，并通过可训练的参数提取分离的季节趋势模式。此外，分层不变性学习器构建了基于变化的过滤器，以实现局部和全局的过滤。

启发：

1. 感兴趣看了看因果学习和AI模型最新的研究

文献3

题目：**Remote sensing of diverse urban environments: From the single city to multiple cities**

作者：Gang Chen, Yuyu Zhou, James. Voogt, Eleanor C. Stokes

出处：RSE 2024

方法：

本文对多城市遥感的科学现状进行了全面的 review，旨在促进这一新兴领域的快速发展，以应对全球可持续性挑战，并支持新学科——城市可持续性科学 (USS) 所需的知识发展。本文通过对城市遥感的 8 个关键研究领域 (即土地利用和土地覆盖及其变化、城市垂直结构、城市热岛、危害、能源利用和排放、空气质量、碳预算和绿地) 的综合综述，对开展多城市研究的基本原理、城市选择标准、社会应用、城市发展前景和城市发展前景等方面进行了深入探讨。以及扩大多城市遥感评估范围的机会和未来方向。

多城市遥感定义：跨越两个或多个不同地理模式的城市的城市的研究，并可以通过高度概括的知识或见解促进对区域或全球范围内城市系统的理解。

多城市研究的未来愿景：

- 用于数据采集的新型传感器系统
- 开放的遥感数据 (如 MODIS、Landsat 和 Sentinel)
- 智能数据处理和分析系统
- 整合其他专业领域的知识，创建一个新的城市科学/选择城市样本的指南：

(i) 创造一种超越单一城市地区的新理论

(ii) 将城市地区作为符合地方、城市、区域和国家空间尺度尺度规律的社会、经济、基础设施和空间复杂系统，并参与在多个尺度上塑造城市化进程的制度政策，以限制意外后果

(iii) 确定可能受益于类似可持续性战略的“对等城市”群体，以有效地扩大行动

(iv) 能够总体检查城市化对地球的影响，例如物种灭绝，排放产生，农业用地流失

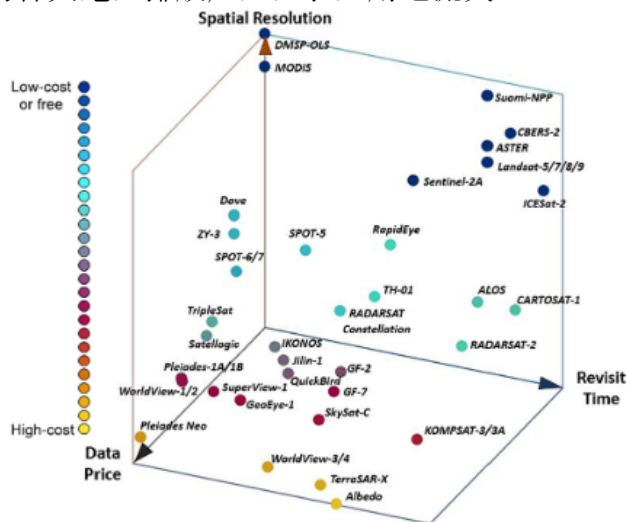
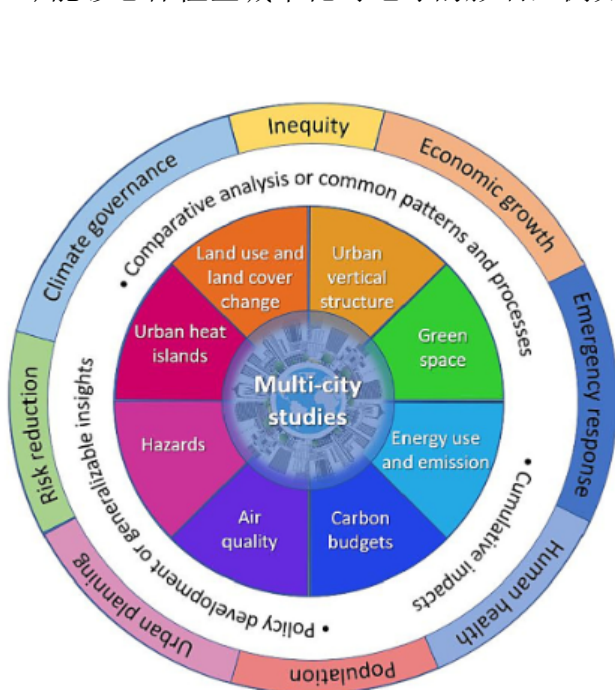


Fig. 3. A 3D cube with axes of spatial resolution (fine to coarse), data price (low to high) and revisit time (short to long), which includes 35 sample satellite sensors for multi-city studies. The blue-to-yellow colour scheme shows the increase of data price from low/free to high. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

启发:

1. 综述主要还是提出了**多城市遥感的新概念**: 跨越两个或多个不同地理模式的城市的研 究, 并可以通过高度概括的知识或见解促进对区域或全球范围内城市系统的理解。
2. 目前研究对大城市的评估存在明显的偏见, 特别是在地理上强调位于中国、欧洲和北美的城市

本周工作

1. 阅读文献
2. 大数据集的用户访问数据用 scala 和 spark 集群语言处理完了
但目前比较麻烦的是, 存在一些用户的访问比较稀疏, 例如用户 A 两次访问记录之间相差 5 天, 目前在做数据集访问记录的统计筛选, 看这类用户大概有多少, 在建立 graph 的时候这种访问可能需要剔除

下周计划

1. 建立大数据集的 graph, 再抽取一个子集作为小数据集跑 graphormer