



BIGSCITY时空数据挖掘研讨会

以不变应万变： 面向时空数据的不变关联学习



数据智能实验室
Data Intelligence Lab

中国科学技术大学 周正阳

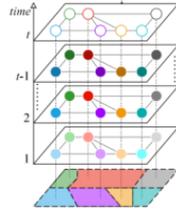
2023.06.16



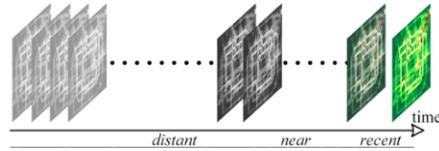
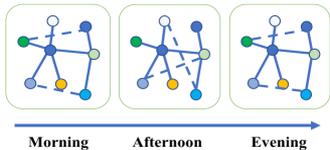
时空学习研究背景

➤ 时空数据一般特性和常见方法

1. 路网传播特性

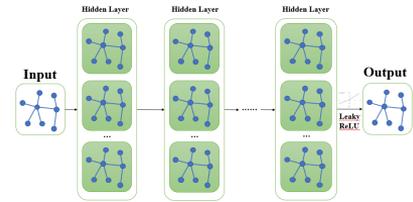
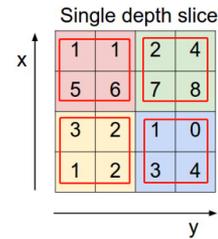


2. 区域之间的动态空间关联

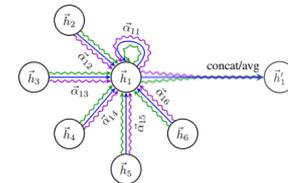


3. 时间序列的趋势性、连续性、周期性

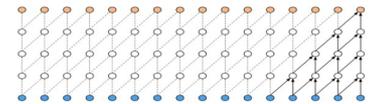
CNN、GCN



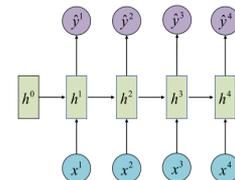
注意力机制



时变模式抽取



RNN、TCN

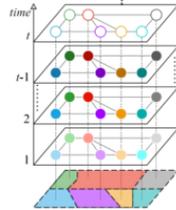




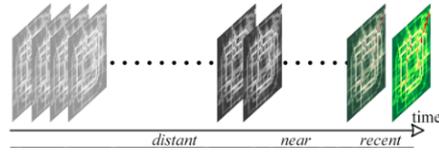
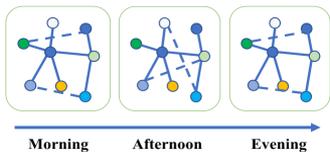
时空学习研究背景

➤ 时空数据一般特性和常见方法

1. 路网传播特性



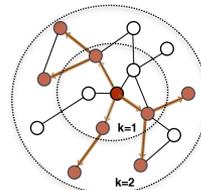
2. 区域之间的动态空间关联



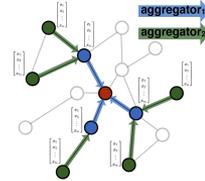
3. 时间序列的趋势性、连续性、周期性

GCN

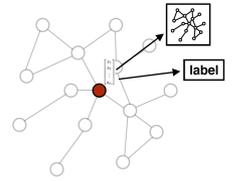
$$Y = f(X, A) = \text{Softmax}(\hat{A} \text{ReLU}(\hat{A}XW^0)W^1)$$



1. Sample neighborhood

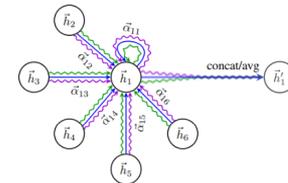


2. Aggregate feature information from neighbors

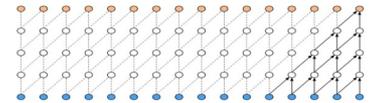


3. Predict graph context and label using aggregated information

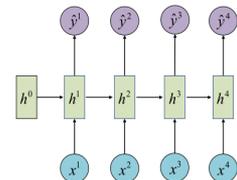
注意力机制



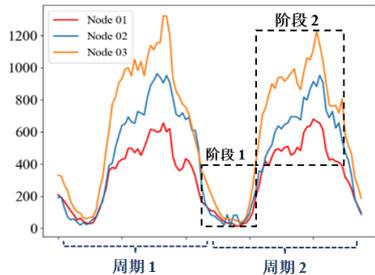
时变模式抽取



RNN、TCN



时空数据的分布偏移问题



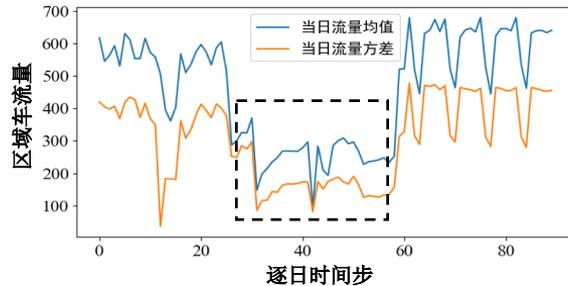
不同阶段展示出不同的模式

自然的



空间结构微调

人造的



城市空间结构微调导致的分布时序偏移

当数据波动与偏移由量变引起质变

$$(\mu_1, \sigma_1) \rightarrow (\mu_2, \sigma_2)$$

$$\mu_2 \gg \mu_1 \downarrow$$

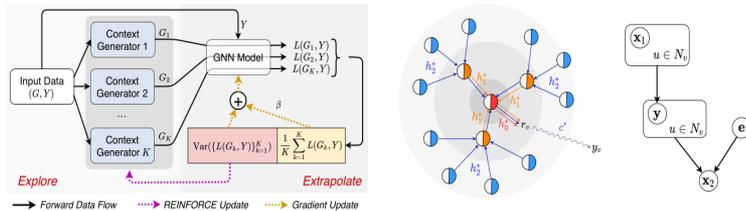
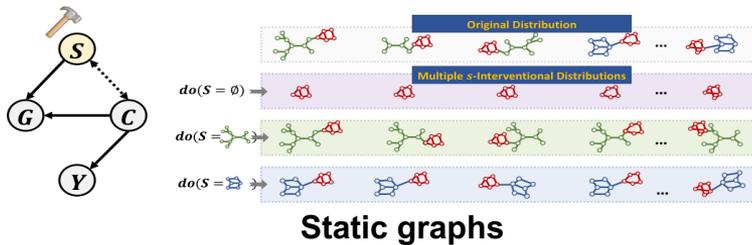
误差放大



PROPOSITION 1. Let the graph-based ST learner f trained without considering invariant relations, then the upper bound of empirical risk under environment e_0 would be $\varepsilon_0 \sim \frac{2(1-p_i)d_i\mu_0 w_i^2}{1+d_i}$ that is irreducible, where w_i^s is the weight for causal neighbor aggregation. When f is transferred to OOD test set $N(\mu_q, \sigma_q|e_q)$ satisfying $\mu_q = q\mu_0$ where $q \in \mathbb{N}^+$. The OOD risks are amplified to $\varepsilon_q \sim \frac{2(1-p_i)d_i q \mu_0 (\mu_w \pm 3\sigma_{w_s})}{1+d_i}$, where μ_w and σ_{w_s} are the expectation and variance of learnable w_s .

非不变学习的误差与观测期望和方差均相关。当 μ_2 远大于 μ_1 且 $\mu_2 = q\mu_1$ 时，非不变学习的误差上界将被大约放大 q 倍，导致较差的泛化性能。

➤ 现有工作&挑战

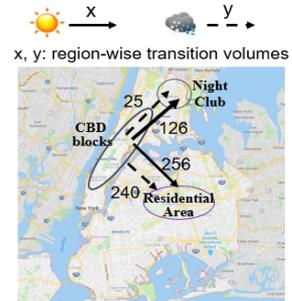
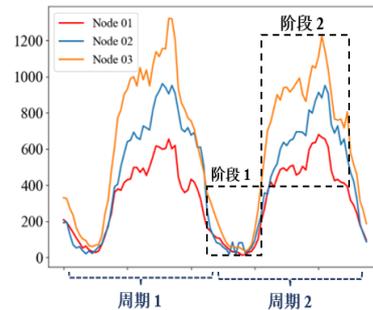


$$\min_{\omega, \Phi} \sum \mathcal{R}^e(\omega, \Phi) + \lambda \text{Var}(\mathcal{R}^e(\omega, \Phi))$$

➤ Subgraph-based invariance extraction

➤ Invariant risk minimization

- [1] Wu Y, Wang X, Zhang A, et al. Discovering Invariant Rationales for Graph Neural Networks[C]//International Conference on Learning Representations 2022.
[2] Wu Q, Zhang H, Yan J, et al. Handling Distribution Shifts on Graphs: An Invariance Perspective[C]//International Conference on Learning Representations 2022.



- 高度动态性: Segment-level heterogeneity
- 回归任务特性: Transition volumes count
- 不可解释性: IRM cannot interpret on how each relation exactly invariant



Capture quantified invariant relations for OOD ST learning

➤ 解决方案总览

PROPOSITION 2. Let the ST learner f^* be trained considering invariant relations, then the empirical risk under any environment $e_i (i = 1, \dots, K)$ can asymptotically converge to 0 with $w_i^c = \frac{1}{p_i}$ where w_i^c is the weight for causal part neighbor aggregations.

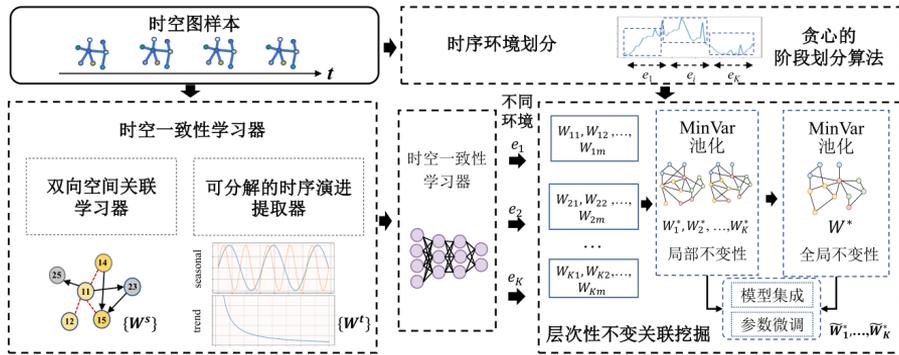


基于不变关联的时空预测分布外泛化模型

抑制虚假邻域关联可使得误差与原始观测期望独立，从而**抑制泛化误差**。

将时间步视作环境

- 时序环境划分
- 时空关联一致性学习器
- 层次性不变关联挖掘





面向时空预测的分布外泛化研究

➤ 时序环境划分

局部环境模式不同



基于贪心算法实现阶段划分

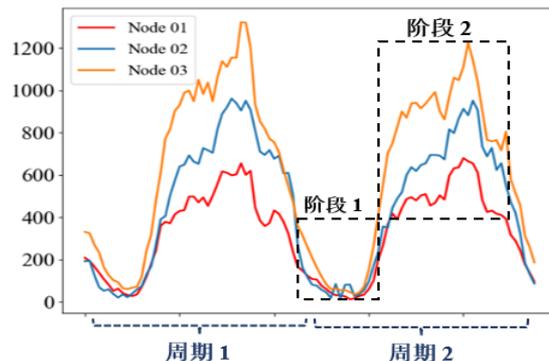
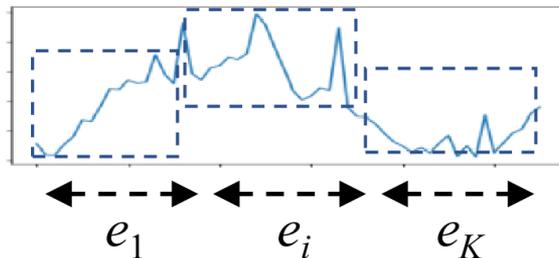
渐近地选取模式最相异的片段

$$\max \sum_{1 \leq p \neq q \leq K} d(\mathbf{D}_p, \mathbf{D}_q), s.t. \forall i, \Delta_1 < |\mathbf{D}_i| < \Delta_2; \sum_{i=1}^K |\mathbf{D}_i| = N$$



$$d(\mathbf{D}_p, \mathbf{D}_q) = \frac{1}{N} \sum_{i=1}^N |\cos(\mathbf{D}_{p_i}, \mathbf{D}_{q_i})|$$

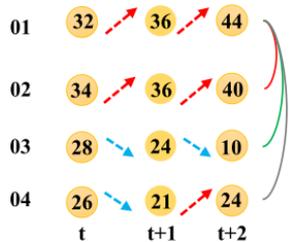
- 保有局部**个性化的不变性**和**双向关联**
- 便于构建**层次化局部-全局不变关联**



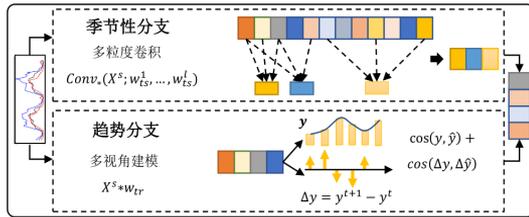


面向时空预测的分布外泛化研究

➤ 时空一致性不变关联学习器



空间一致性学习器



时序解耦一致性学习

$$\mathbf{h}^{t+1} = (\mathbf{A}_0 \mathbf{X} + \mathbf{A}_{in} \mathbf{X}^{-1}) \mathbf{U}$$



- 将时空关联转化为可学习权值
- 可基于权值波动选择稳定关联
- 同时捕获稳定的正向和负向关联

Spatial relation

$$Relation \ r^t(v_i, v_j) = \frac{x_i^{t+1} - \bar{x}_i^t \bar{x}_j^t}{x_j^{t+1} - \bar{x}_j^t \bar{x}_i^t}$$

$$x_i^{t+1} = \bar{x}_i^t + \frac{r_{ij}^t (x_j^{t+1} - \bar{x}_j^t) \bar{x}_i^t}{\bar{x}_j^t}$$

$$E(h_i^{t+1}) = k_0 x_i + \sum_{\substack{v_j \in \mathcal{N}(v_i), \\ s \in [1, |\mathcal{N}(v_i)|]}} k_{s1} x_j + \frac{k_{s2}}{x_j}$$

$$\mathbf{h}^{t+1} = (\mathbf{A}_0 \mathbf{X} + \mathbf{A}_{in} \mathbf{X}^{-1}) \mathbf{U}$$

Temporal dependence

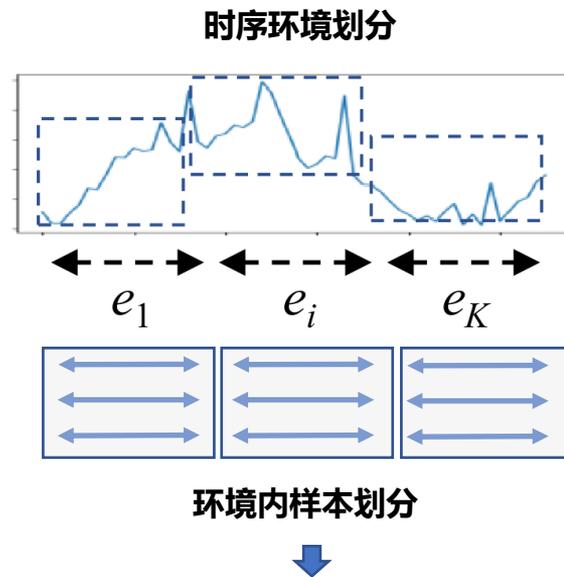
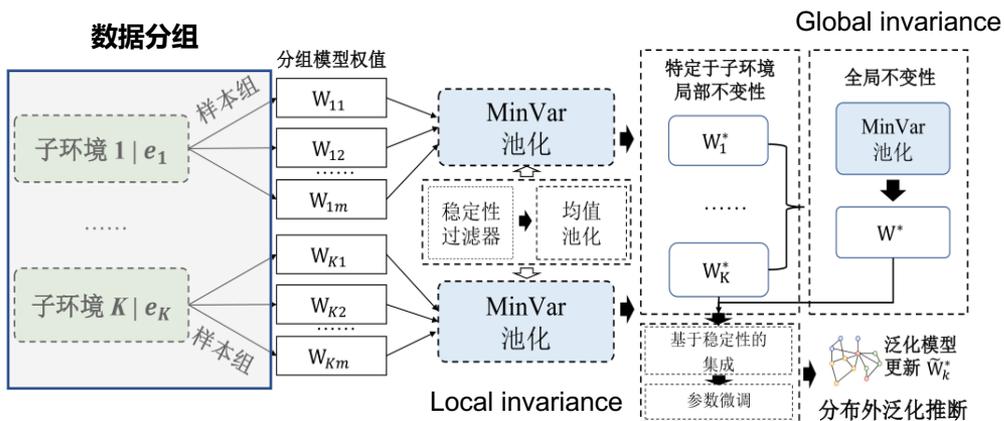
$$d_{c_k} = 1/f_k \quad (k = 1, 2, \dots, l)$$

$$\hat{\mathbf{Y}}_s^{t+1:t+\tau} = \text{TCN}(\mathbf{X}^s; w_{ts}^1, \dots, w_{ts}^l)$$

$$\text{Reg}_{in} = -\min \cos(\hat{\mathbf{Y}}_s, \hat{\mathbf{Y}}_{tr})$$

$$Loss = \text{MAPE}(\hat{\mathbf{Y}}, \mathbf{Y}) + \lambda_0 \text{Loss}_{se} + \lambda_1 \text{Loss}_{tr} + \lambda_2 \text{Reg}_{in}$$

➤ 层次性不变关联挖掘

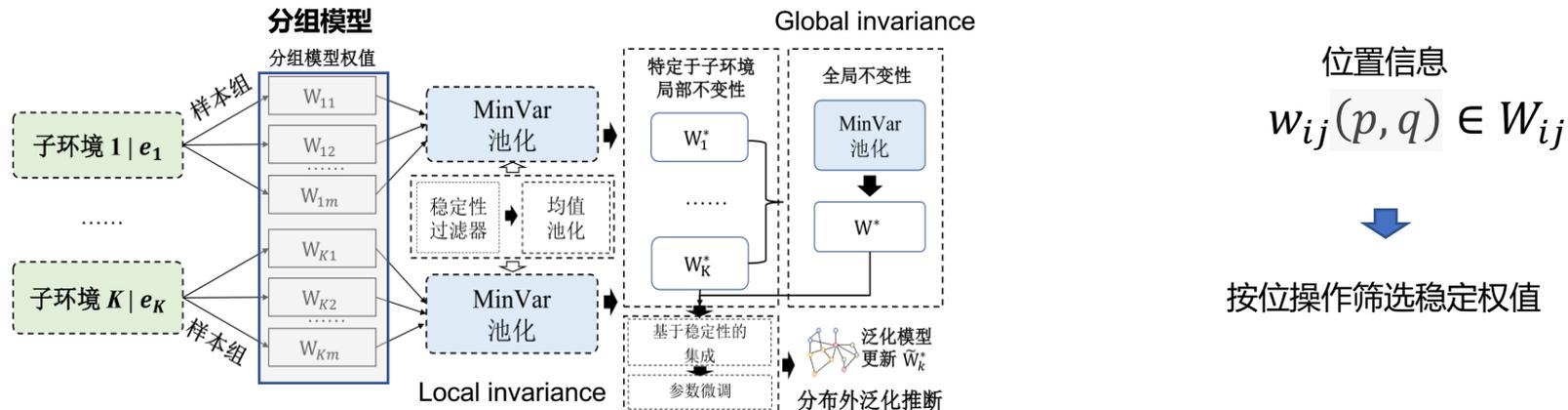


每个环境m组样本，分别训练m个模型

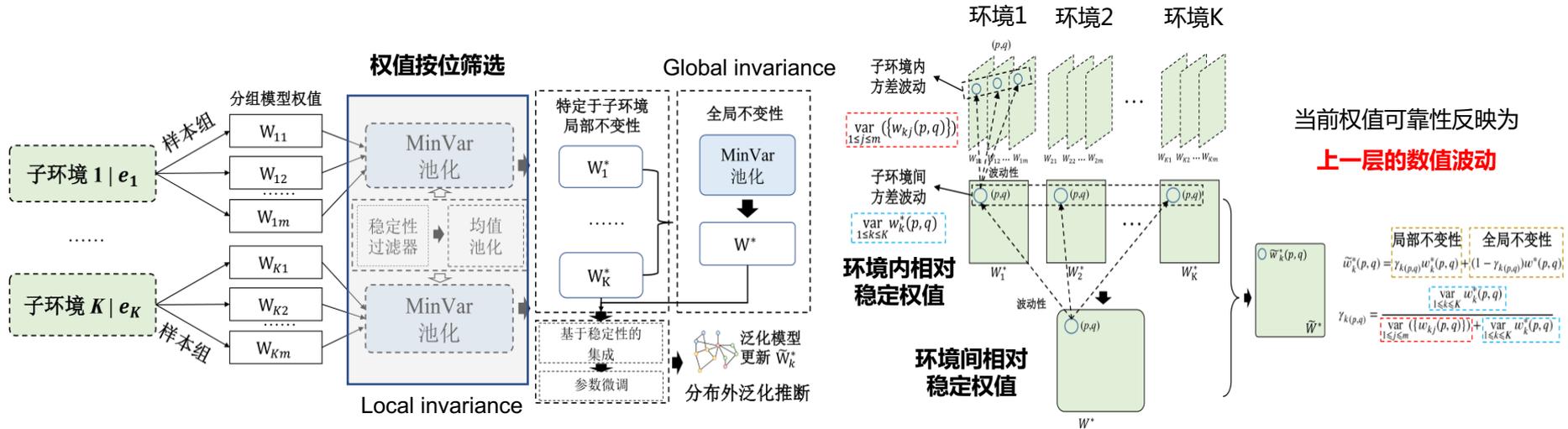


面向时空预测的分布外泛化研究

➤ 层次性不变关联挖掘



➤ 层次性不变关联挖掘

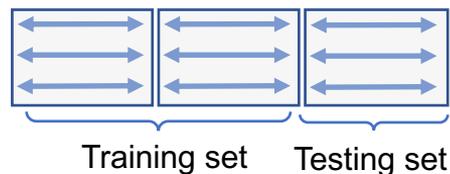
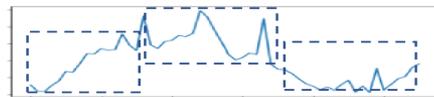


$$(p, q) = \arg \min_{\text{Min-r}\%} (\text{Var}_{1 \leq j \leq m} (\{w_{kj}(p, q)\})) \rightarrow w_k^*(p, q) = \text{Avg}_{1 \leq j \leq m} (\{w_{kj}(p, q)\})$$

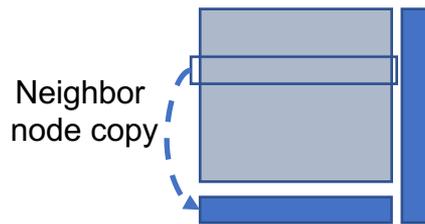
基于稳定性的权值过滤与集成

实验：数值分析

- ✓ **时序偏移**：基于segment划分训练-测试集，模拟随时间变化产生分布偏移现象
- ✓ **新节点归纳推断**：训练阶段mask部分节点，并在测试阶段重新引入，基于Nodecopy策略补充邻接阵
- ✓ **人造噪声**：主动注入噪声以模拟分布波动

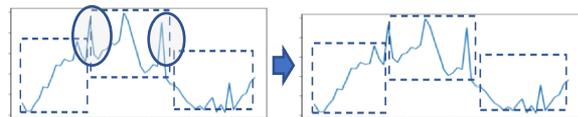


(a) 时序偏移



(b) 新节点归纳

Noise injection



(c) 噪声注入

数据集统计信息

数据集	节点数	时间步数量	时间跨度	间隔长度	周期性内均值/ 周期内标准差
SIP	108	25,920	01/01/2017- 03/31/2017	5min	72.17/55.45
Metr-LA	207	34,272	03/01/2012- 06/30/2012	5min	27.45/30.25
KnowAir	184	11,688	01/01/2015- 12/31/2018	3h	52.69/61.60
Electricity	321	26,304	01/01/2012- 12/31/2014	15min	2538.79/820.92



面向时空预测的分布外泛化研究

➤ 实验：数值分析

- ✓ **时序偏移**：基于segment划分训练-测试集，模拟随时间变化产生分布偏移现象
- ✓ **新节点归纳推断**：训练阶段mask部分节点，并在测试阶段重新引入，基于Nodecopy策略补充邻接阵
- ✓ **人造噪声**：主动注入噪声以模拟分布波动

数据集统计信息

数据集	节点数	时间步数量	时间跨度	间隔长度	周期性内均值/ 周期内标准差
SIP	108	25,920	01/01/2017- 03/31/2017	5min	72.17/55.45
Metr-LA	207	34,272	03/01/2012- 06/30/2012	5min	27.45/30.25
KnowAir	184	11,688	01/01/2015- 12/31/2018	3h	52.69/61.60
Electricity	321	26,304	01/01/2012- 12/31/2014	15min	2538.79/820.92

周期性相关设定与子环境划分后平均方差

数据集	SIP	Metr-LA	KnowAir	Electricity
时间步	25,920	34,272	11,688	26,304
时间间隔	5 min	5 min	3 h	15 min
周期	daily	daily	weekly	daily
周期步数	288	288	56	96
子环境划分个数 K	6	6	6	6
子环境划分后阶段内平均方差	25.89	10.32	34.85	445.24

模型超参设置

参数	具体数值
CauSTG 主干网络	GraphWaveNet (GWN)
训练样本分割比例	训练/验证/测试集比例: 2/1/1
子环境划分个数 K	6
子环境内的模型个数 m	4
时序模式提取任务平衡系数 $(\lambda_0, \lambda_1, \lambda_2)$	(0.5, 0.5, 0.2)
学习率	1e-3
不变性比例阈值 $r\%$	(60%, 55%, 55%, 50%)
TCN 核的个数 l	(5, 5, 6, 6)
TCN 核的维度	(12, 6, 3)
新节点推断中新节点比例	5%
GNN 的隐层维度 d_m	64
优化器	Adam



面向时空预测的分布外泛化研究

实验：数值分析

性能比较：SIP and Metr-LA

	SIP			Metr-LA		
	时序偏移	新节点 归纳推理	人造噪声	时序偏移	新节点 归纳推理	人造噪声
STGCN	22.75%	26.72%	23.36%	12.62%	15.13%	13.53%
MTGNN	20.09%	23.74%	20.70%	10.05%	12.56%	11.25%
GWN	20.13%	23.65%	20.84%	10.01%	12.52%	11.17%
DCRNN	21.17%	24.64%	21.88%	10.50%	13.01%	11.41%
ASTGNN	22.31%	25.87%	22.92%	10.04%	12.55%	10.99%
AdaRNN	21.22%	24.78%	22.79%	10.14%	13.08%	11.58%
MTGNN+IRM	20.21%	23.86%	20.52%*	10.02%	12.53%	11.03%
GWN+IRM	20.01%*	23.56%*	20.64%	9.94%*	12.45%*	10.95%*
CauSTG	19.91%	23.03%	20.35%	9.75%	12.34%	10.64%
Beyond non-inv	0.90%	2.62%	1.69%	2.60%	1.44%	3.18%
Beyond inv	0.50%	2.25%	0.83%	1.91%	0.88%	2.83%

消融实验

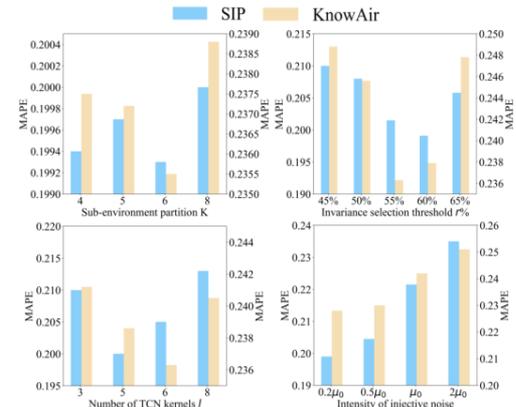
消融变体	SIP	Metr-LA	KnowAir	Electricity
CauSTG-Adj	21.10%	11.60%	26.14%	2.24%
CauSTG-GRU	21.62%	10.55%	25.17%	2.30%
CauSTG-NoHier	23.26%	13.42%	26.68%	2.84%
CauSTG	19.91%	9.75%	23.63%	1.89%

性能比较：KnowAir and Electricity

	KnowAir			Electricity		
	时序偏移	新节点 归纳推理	人造噪声	时序偏移	新节点 归纳推理	人造噪声
STGCN	31.71%	42.87%	33.94%	2.65%	4.93%	3.93%
MTGNN	24.06%	36.22%	29.59%	2.12%	4.41%	3.65%
GWN	24.13%	36.21%	29.53%	2.08%	4.34%	3.51%
DCRNN	25.17%	36.23%	30.43%	2.31%	4.68%	3.83%
ASTGNN	26.31%	37.43%	31.27%	2.33%	4.71%	3.79%
AdaRNN	24.60%	36.47%	30.76%	2.10%	4.45%	3.88%
MTGNN+IRM	24.03%	35.14%	29.46%	2.05%	4.21%	3.28%*
GWN+IRM	24.01%*	35.12%*	29.34%*	2.04%*	4.13%*	3.33%
CauSTG	23.63%	34.32%	28.95%	1.89%	3.89%	3.15%
Beyond non-inv	1.79%	5.22%	1.96%	9.13%	10.37%	10.26%
Beyond inv	1.58%	2.28%	1.33%	7.35%	5.81%	3.96%

超参数调整

- ✓ Tradeoff between efficiency and performance
- ✓ Tradeoff between variance and invariance

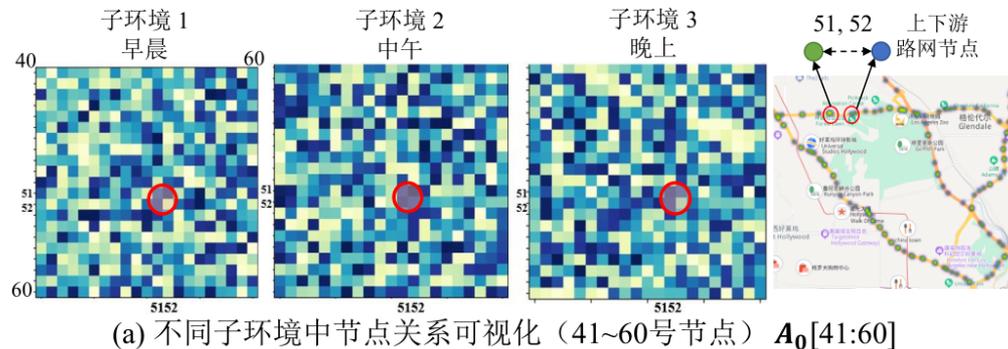




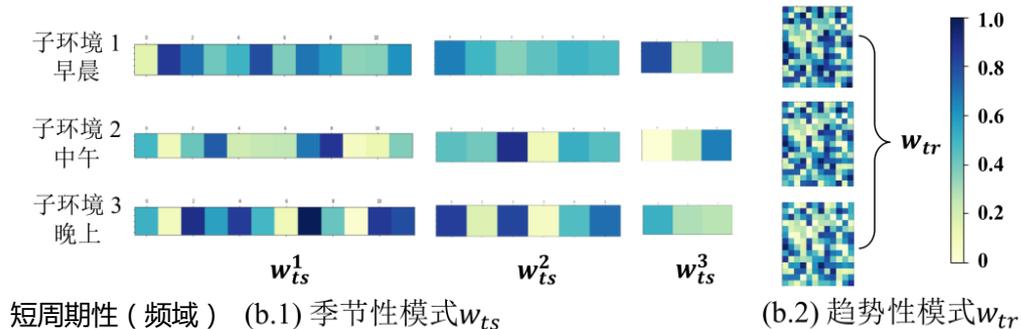
面向时空预测的分布外泛化研究

实验：CauSTG模型权值可视化

空间关联



时序依赖



(b) 不同子环境的时序模式可视化

模型所捕获的稳定关联具有一定的物理解释性

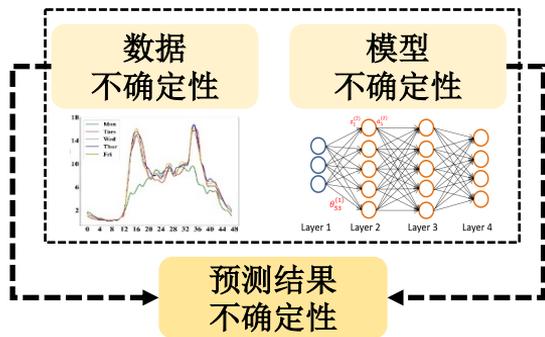


延伸：面向时空图学习的高效可靠模型设计

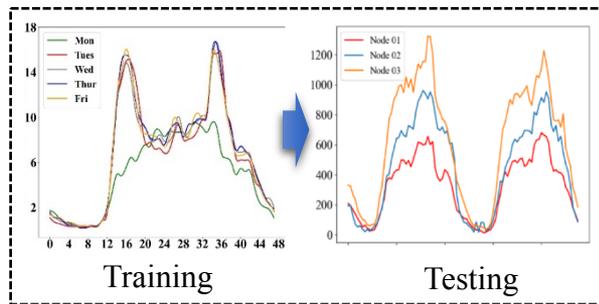
动态性、异质性 → 模型部署挑战 →

数据波动适应性 $\begin{cases} \rightarrow \text{(a) 不确定性量化} \\ \rightarrow \text{(b) 分布外泛化} \end{cases}$

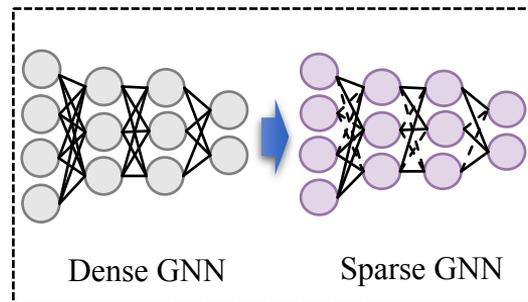
边缘设备适配性 $\rightarrow \text{(c) 模型轻量化}$



(a) 不确定性量化问题



(b) 分布外泛化问题



(c) 模型剪枝与轻量化问题

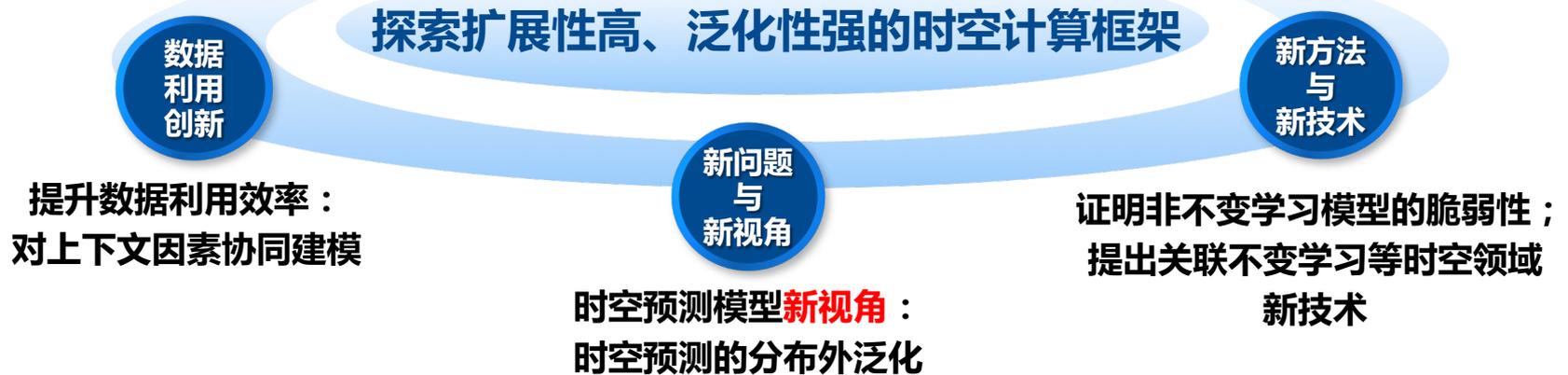
相辅相成





总结与评价

➤ 主要创新点





主要内容

未来研究工作展望

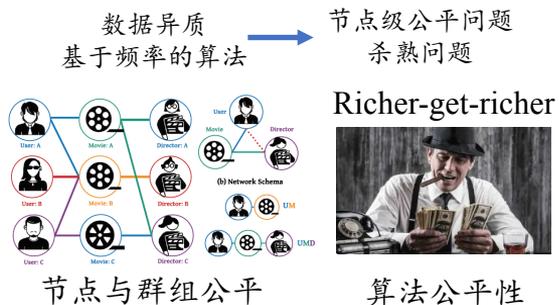


拟开展的研究工作

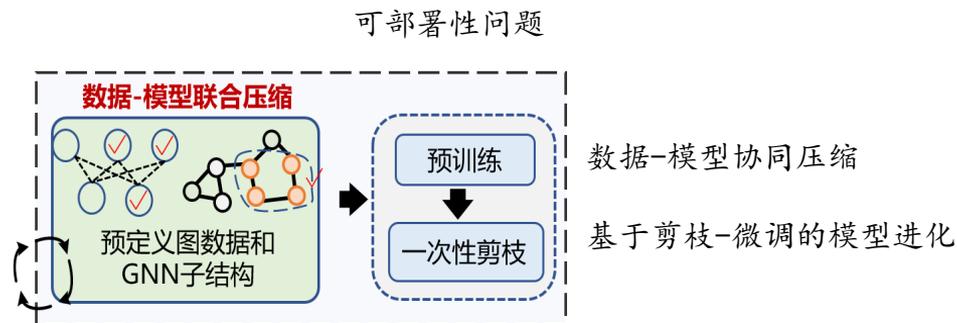
时空数据的预测新理论

挖掘数据新特性 $\xrightarrow{\text{构建}}$ 预测新理论

时空图预测公平性理论



时空数据模型压缩与训练策略



模型公平性、推理高效性与模型可部署性



BIGSCITY时空数据挖掘研讨会

欢迎交流!



数据智能实验室

Data Intelligence Lab



个人主页

中国科大 周正阳

联系方式 : zzy0929@mail.ustc.edu.cn

<https://github.com/zzyy0929/KDD23-CauSTG>



与我联系