

Localised Adaptive Spatial-Temporal Graph Neural Networks

· 时空数据挖掘技术研讨会 ·

韩程凯

□ 我们是否以及在多大程度上可以对时空图模型进行局部化？

- 我们多大程度上需要图结构的信息呢？能否在不影响效果的前提下，减少图结构信息的引入，以此提升模型效率呢？
- 讨论范围限制在自适应时空图神经网络（ASTGNNs）。
- ASTGNNs的局部化是通过稀疏化邻接矩阵来实现的，即剪枝空间图中的边。

□ 研究时空图模型局部化的原因：

- 对数据中的空间和时间依赖性有更深入的理解。
- 资源高效的ASTGNNs设计。
- 时空图模型的分布式部署。

□ 问题定义

$$\{X_t, \dots, X_{t+\mathcal{H}}\} = \mathcal{F}(X_{t-\mathcal{T}}, \dots, X_{t-1}; \theta, \mathcal{G})$$

□ 空间网络建模

- 经典STGNNs中的GCNs:

$$Z_t = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A}_t \tilde{D}^{-\frac{1}{2}} X_t W \right)$$

- [NIPS20]AGCRN, 将规范化自适应邻接矩阵与节点自适应参数学习相结合来增强GCN层 (NAPL-AGCN) :

$$A_{adp} = \text{SoftMax} \left(\text{ReLU} \left(\mathbf{E} \mathbf{E}^T \right) \right)$$

$$Z_t = \sigma \left(A_{adp} X_t \mathbf{E} W_{\mathcal{G}} \right)$$

- AGFormer: 用NAPL-AGCN代替多头自注意力机制中的MLP层, 构建了一个基于Transformer的ASTGNN模型

Adaptive Graph Sparsification

□ 基于NAPL-AGCN的ASTGNNs的训练目标:

$$\mathcal{L}(\theta, \mathbf{A}_{adp}) = \frac{\sum_{\tau \in \mathcal{T}} \sum_{v \in \mathcal{V}} \left\| \mathbf{y}^{(\tau, v)} - \hat{\mathbf{y}}^{(\tau, v)} \right\|_1}{|\mathcal{V}| \times |\mathcal{T}|}$$

□ 引入了一个掩码矩阵 \mathbf{M}_A 来剪枝邻接矩阵 \mathbf{A}_{adp} , 优化以下目标函数:

$$\mathcal{L}_{AGS} = \mathcal{L}(\theta, \mathbf{A}_{adp} \odot \mathbf{M}_A) + \lambda \|\mathbf{M}_A\|_0,$$

$$\|\mathbf{M}_A\|_0 = \sum_{i=1}^N \sum_{j=1}^N m_{(i,j)}, m_{(i,j)} \in \{0, 1\}$$

□ 获得 \mathbf{M}_A 的一种方法是初始化一个可训练的权重矩阵 \mathbf{U} , 用伯努利分布映射到二元门, 但该方法有两个问题:

- 对于大型图来说, 它可能是unscalable的。
- L_0 稀疏度惩罚函数是不可微的。

Adaptive Graph Sparsification

- 对于scalability问题，采用了节点自适应权重学习技术，用节点嵌入E来生成U，以降低计算成本：

$$U = EW_E$$

- 对于不可微的问题，引入了hard concrete分布，该分布是离散分布的连续松弛并且可以近似二元值：

$$z \sim \mathcal{U}(0, 1), s_{(i,j)} = \text{Sig} \left(\log z - \log(1 - z) + \log \left(u_{(i,j)} \right) \right) / \beta$$
$$\bar{s}_{(i,j)} = s_{(i,j)} (\zeta - \gamma) + \gamma, m_{(i,j)} = \min \left(1, \max \left(0, \bar{s}_{(i,j)} \right) \right)$$

Adaptive Graph Sparsification

Algorithm 1: Adaptive Graph Sparsification (AGS)

Input: \mathcal{X} : input data, $\mathcal{F}(\cdot; \theta, \mathbf{A}_{adp})$: Spatial-temporal GNN with initialization self-adaptive adjacency matrix \mathbf{A}_{adp} , N_1 : number of pre-training iterations, N_2 : number of sparsification iterations, s_g : pre-defined sparsity level for graph.

Output: $\mathcal{F}(\cdot; \theta, \mathbf{A}_{adp} \odot \mathbf{M}_A)$

- 1: **while** iteration $i < N_1$ **do**
 - 2: Forward to compute the loss in Eq.(4).
 - 3: Back-propagate to update θ and \mathbf{A}_{adp} .
 - 4: **end while**
 - 5: Obtain pre-trained $\mathcal{F}(\cdot; \theta, \mathbf{A}_{adp})$.
 - 6: Sort entries in \mathbf{A}_{adp} by magnitude in an ascending order then obtain list $K = \{k_i\}_{i=1}^{N \times N}$.
 - 7: **while** iteration $i < N_2$ and $1 - \frac{\|\mathbf{M}_A\|_0}{\|\mathbf{A}_{adp}\|_0} < s_g$ **do**
 - 8: set $\mathbf{M}_A^{(i,j)} = 1$ if $\mathbf{A}_{adp}^{(i,j)} \notin K_{:N^2 s_g}$.
 - 9: Forward to compute the loss in Eq.(5).
 - 10: Back-propagate to update θ and $\mathbf{A}_{adp} \odot \mathbf{M}_A$.
 - 11: **end while**
-

□ 用AGS剪枝邻接矩阵显著降低了用于推理的ASTGNNs的计算复杂度:

- 未剪枝NAPL-AGCN层的推断时间复杂度:

$$\mathcal{O}(N^2 d + LNTF^2 + LT\|\mathbf{A}_{adp}\|_0 + NdF)$$

- 稀疏化NAPL-AGCN层的推断时间复杂度:

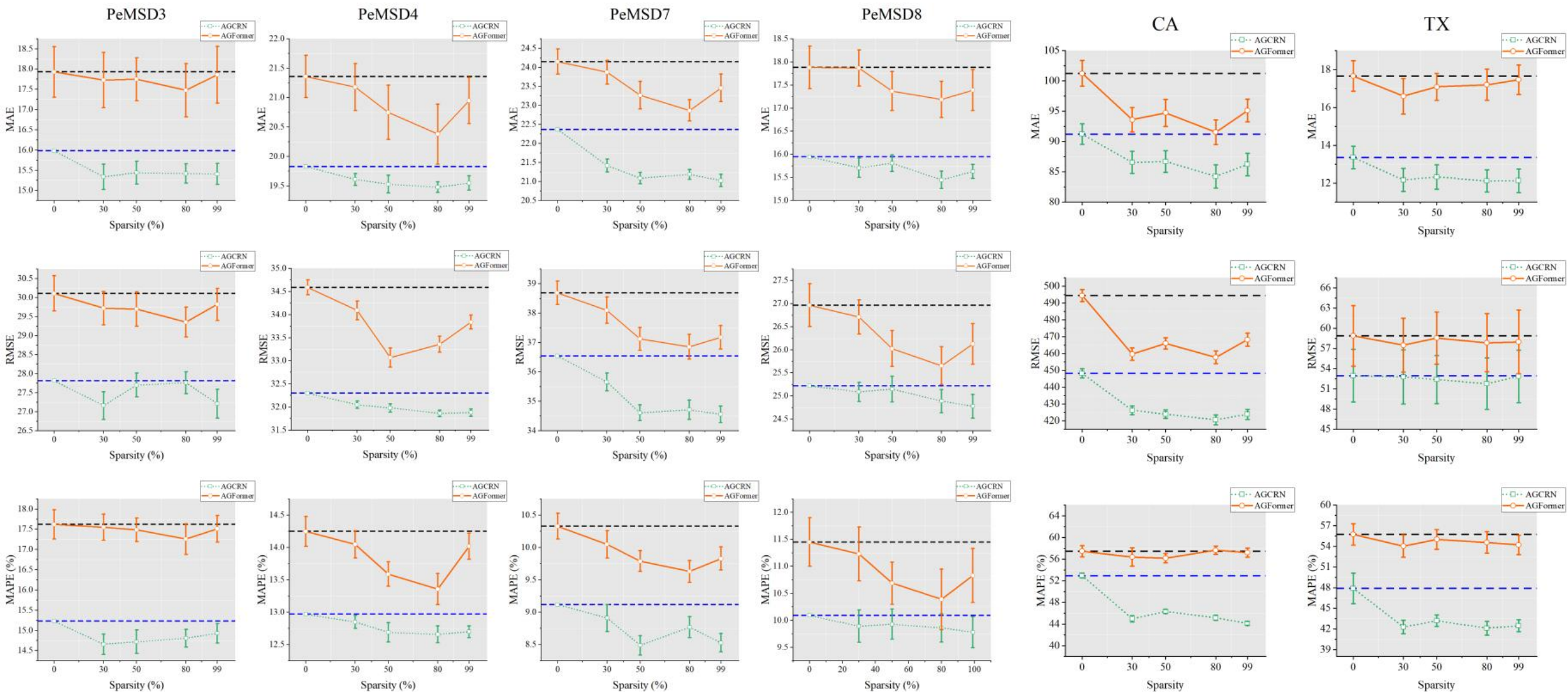
$$\mathcal{O}(N^2 d + LNTF^2 + LT\|\mathbf{A}_{adp} \odot \mathbf{M}_A\|_0 + NdF)$$

Datasets and Configurations

Datasets	#Nodes	Range
PeMSD3	358	09/01/2018 - 30/11/2018
PeMSD4	307	01/01/2018 - 28/02/2018
PeMSD7	883	01/07/2017 - 31/08/2017
PeMSD8	170	01/07/2016 - 31/08/2016
Bytom	100	27/07/2017 - 07/05/2018
Decentral	100	14/10/2017 - 07/05/2018
Golem	100	18/02/2017 - 07/05/2018
CA	55	01/02/2020 - 31/12/2020
TX	251	01/02/2020 - 31/12/2020

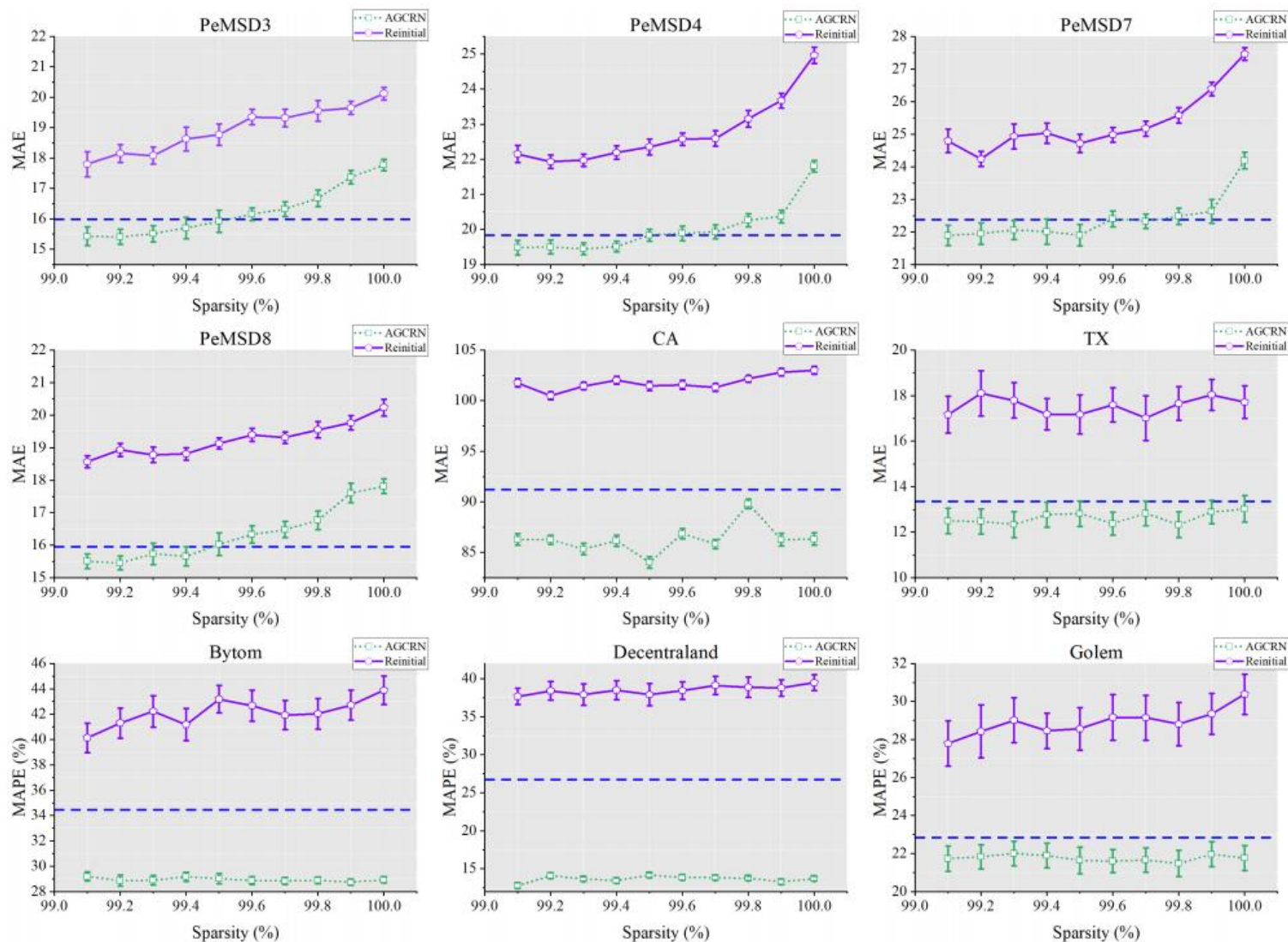
Main Experimental Results

□ 结论1: AGCRN和AGFormer的局部化是可能的。



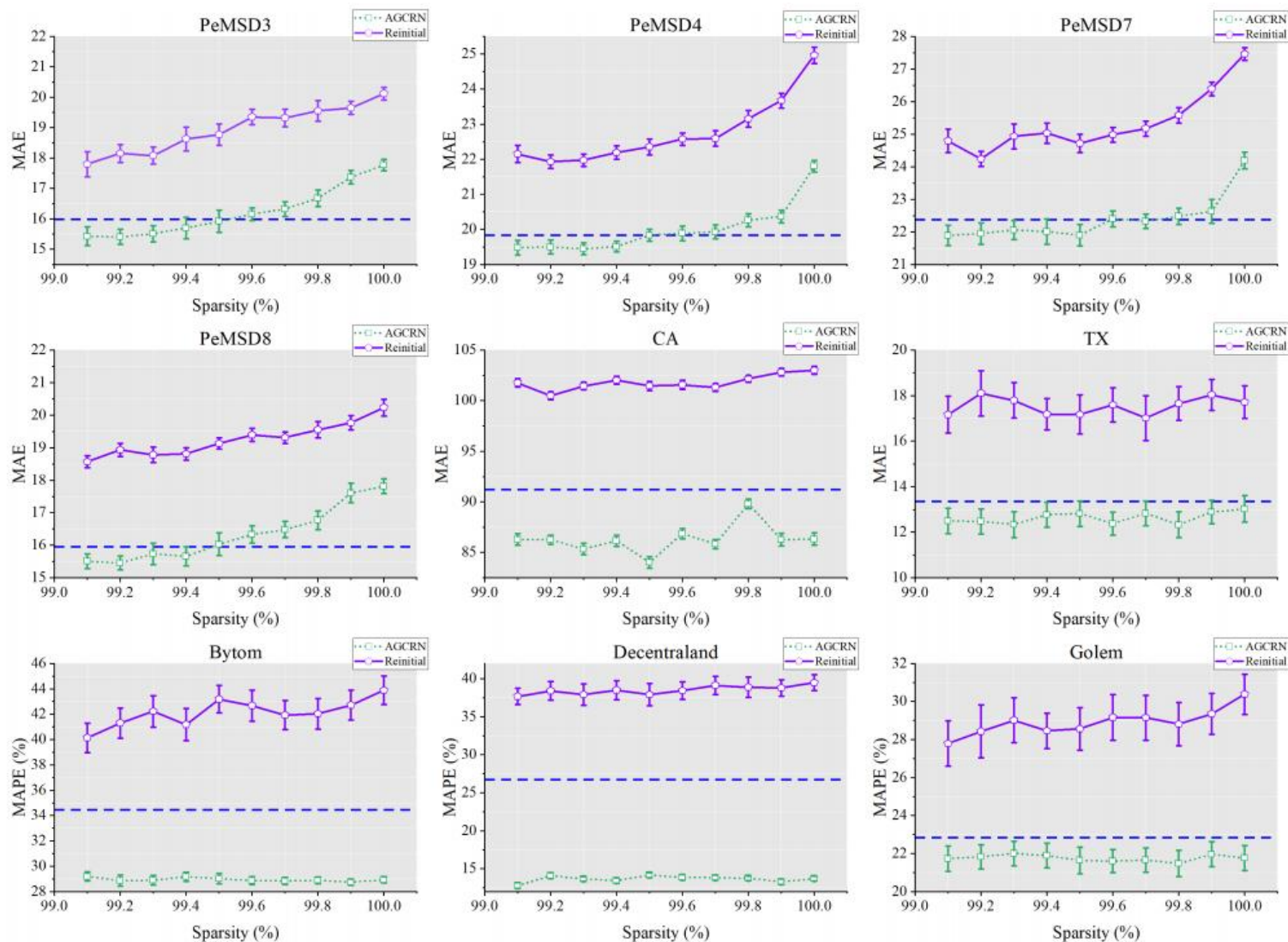
Main Experimental Results

□ 结论2: AGCRN的完全局部化仍然是可行的。



Main Experimental Results

□ 结论3: 如果没有密集的空间图, 局部化AGCRNs就无法重新学习。



Main Experimental Results

□ 基于实验结论，提出以下假设：

- **在许多时空数据集中，由空间相关性提供的信息主要包括在由时间相关性所提供的信息中。因此，可以安全地忽略空间相关性进行推理，而不会显著损失准确性。**
- **尽管包含在空间和时间依赖关系中的信息重叠，但这种重叠提供了正确训练时空图模型所需的重要冗余。因此，在训练过程中不能忽略空间相关性。**

Impact on Resource Efficiency

- **AGCRN和AGFormer的局部化有效地减少了推理所需的计算量。**与AGFormer相比，AGCRN上的加速度更为显著，因为AGFormer所需的总计算量的很大一部分用于其时间模块（Transformer层），而AGCRN使用的时间模块（RNN层）轻得多。

Table 2: Computation cost during inference on original and 99%-localised AGCRNs and AGformers. The amount of computation is measured in MFLOPs, and acceleration factors are calculated in the round brackets.

Methods	Computation Cost for Inference (MFLOPs)								
	PeMSD3	PeMSD4	PeMSD7	PeMSD8	Decentraland	Bytom	Golem	CA	TX
Original AGCRN	400.26	188.59	1131.41	153.06	8.58	8.58	8.58	161.42	850.29
Original AGFormer	122.01	99.45	453.89	47.39	15.02	15.02	15.02	21.50	266.97
Localised AGCRN	253.33(↑1.6×)	80.55(↑2.3×)	237.56(↑4.8×)	119.93(↑1.3×)	2.82(↑3.0×)	2.82(↑3.0×)	2.82(↑3.0×)	145.50(↑1.1×)	706.21(↑1.2×)
Localised AGFormer	80.13(↑1.5×)	68.64(↑1.4×)	199.17(↑2.3×)	37.95(↑1.3×)	11.75(↑1.3×)	1.75(↑1.3×)	11.75(↑1.3×)	19.56(↑1.1×)	226.43(↑1.1×)

Localised AGCRNs vs. Other Non-Localised ASTGNNs

- **高达99%的局部化能够略微提高测试精度。** 这种改进是由稀疏空间图的正则化效应引起的，这可能表明未局部化AGCRN和AGFormer都在一定程度上遭受过拟合。
- **即使与SOTA的架构相比，局部化AGCRN也可以提供有竞争力的推断性能。** 因此，假设时间模块经过适当训练，不同的空间模块可能不会在推理性能上产生显著差异。

Table 3: Performance of 99%-localised AGCRNs compared with other non-localised ASTGNN architectures on transportation datasets.

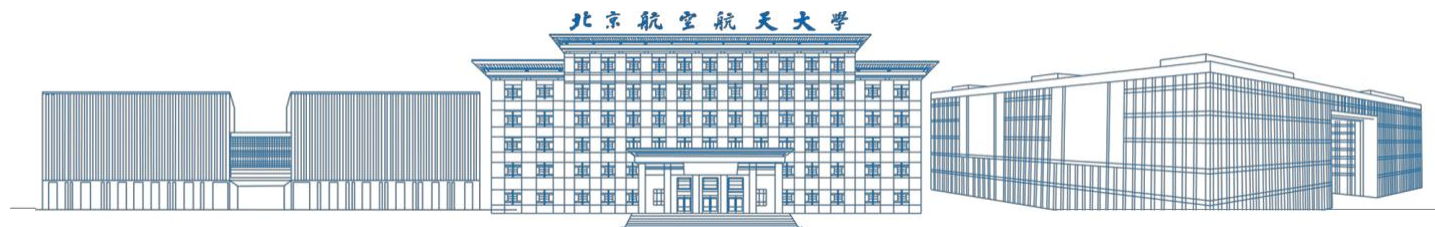
Methods	Datasets	PeMSD3			PeMSD4			PeMSD7			PeMSD8			Average		
	Metrics	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
AGCRN		15.98	28.25	15.23%	19.83	32.30	12.97%	22.37	36.55	9.12%	15.95	25.22	10.09%	18.53	30.58	11.85%
Z-GCNETs		16.64	28.15	16.39%	19.50	31.61	12.78%	21.77	35.17	9.25%	15.76	25.11	10.01%	18.42	30.01	12.11%
STG-NCDE		15.57	27.09	15.06%	19.21	31.09	12.76%	20.53	33.84	8.80%	15.45	24.81	9.92%	17.69	29.21	11.64%
TAMP-S2GCNets		16.03	28.28	15.37%	19.58	31.64	13.22%	22.16	36.24	9.20%	16.17	25.75	10.18%	18.49	30.48	11.99%
Localised AGCRN		15.41	27.21	14.93%	19.55	31.88	12.70%	21.03	34.56	8.53%	15.63	24.78	9.78%	17.91	29.61	11.49%

Localisation of Non-temporal Graphs

- **所有局部化的非时间图模型测试精度显著下降。** 在没有时间依赖性的情况下，空间依赖性提供的信息在推理过程中发挥着重要作用，因此不能通过局部化来忽略。

Table 6: Classification accuracy (%) of localised GCN and GAT on citation graph datasets.

Sparsity(%)	Cora		Citeseer		PubMed	
	GCN	GAT	GCN	GAT	GCN	GAT
0%	80.20	82.10	69.40	72.52	78.90	79.00
30%	80.35	83.17	69.23	72.31	79.14	79.23
50%	72.73	75.40	69.37	72.70	78.82	79.31
80%	65.19	70.81	58.47	63.18	68.37	77.03
100%	56.22	63.29	53.13	57.50	61.02	64.25



感 谢 观 看

· 时空数据挖掘技术研讨会 ·

2023年9月7日