

Networked Time Series Imputation via Position-aware Graph Enhanced Variational Autoencoders

KDD 2023



Time series imputation

多元时序插值问题是近年来研究的热点问题，该问题旨在给定一组**不完整、不规则、**或者有缺失时间点的多元时序数据时，通过某种插值方法来估计在缺失时间点上各个变量的取值。

该问题应用领域广泛，例如气象观测数据处理和预测模型建立、金融市场数据处理和风险管理、以及生态学、医学等，具有很高的研究价值

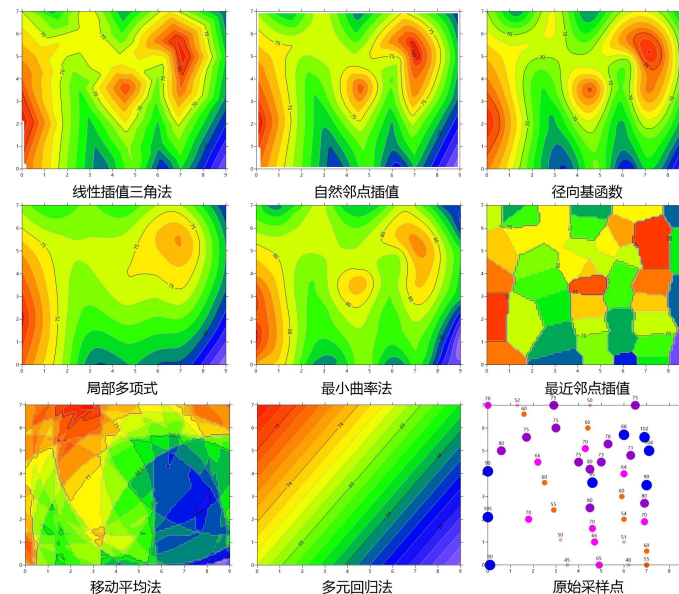
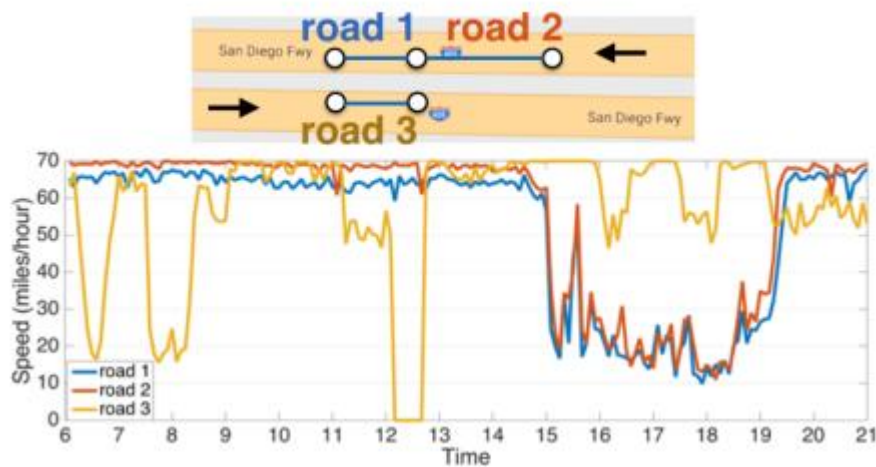


图1. 时序插值的部分应用场景

Time series imputation

深度学习中目前主流插值方法可以分为两大类，包括：

(1)深度循环网络及生成模型，主要关注时间序列特征；

- 将时序数据切分为输入序列和目标序列，用于训练深度循环模型。
- 利用已有数据对深度循环模型进行训练，以学习时序数据的模式和规律，而后使用已训练的深度循环模型作为生成模型的一部分。
- 对于缺失数据点，使用已经训练好的深度循环模型进行预测，得到插值的估计值。

(2)基于图神经网络(gnn)的模型，利用MTS固有图结构的拓扑信息作为关系归纳偏差进行imputation。

- 时序数据表示为图结构，其中每个变量被视为图中的一个节点，而时间上的关联被视为图中的边
- 选择合适的图神经网络模型，将输入定义为图结构的时序数据，即节点特征为各个变量在不同时间点的取值，边特征为时序关联的权重。
- 利用已经训练好的图神经网络模型对整个图进行前向传播，预测缺失节点的特征。

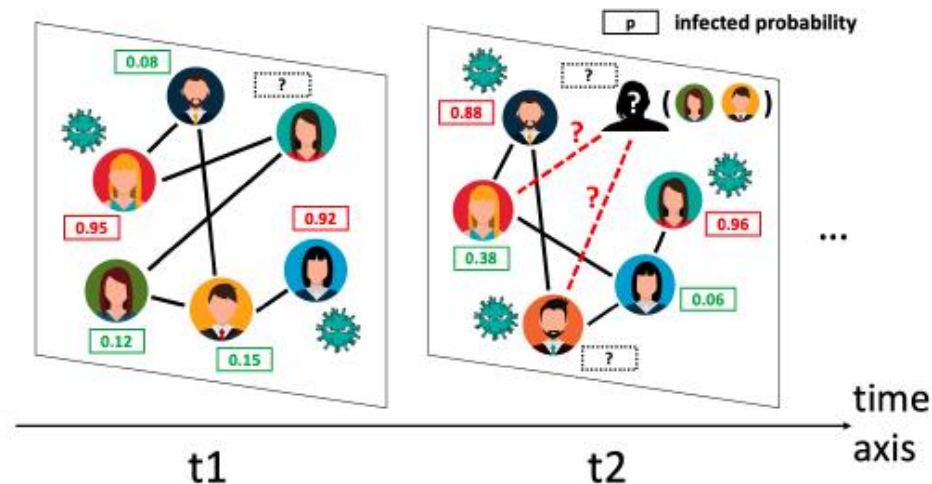
Challenge

传统图神经网络，如图卷积网络GCN，图注意力网络GAT等，图底层拓扑结构固定，然而在现实场景中，多元时间序列的图结构可能会不断变化，如流行病学研究，由于个体的移动，其图结构将为动态。已有图网络不能很好的捕捉这样的特质，该类数据的插值将更加困难。

把这样的时间序列数据称为网络时间序列（NTS），其插值挑战主要有以下两点：

- 时间序列数据背后的图是不断演变的，并且包含缺失的边。因此，算法应该捕捉图的动态信息，同时能够恢复丢失的结构。
- 节点特征时间序列也存在缺失值，这也要求模型解决一般的MTS插值问题（节点与边的缺失）

图2: Covid-19大流行期间的
NTS交互实例



METHODOLOGY

PoGeVon (Position-awareGraph Enhanced Variational Autoencoders)

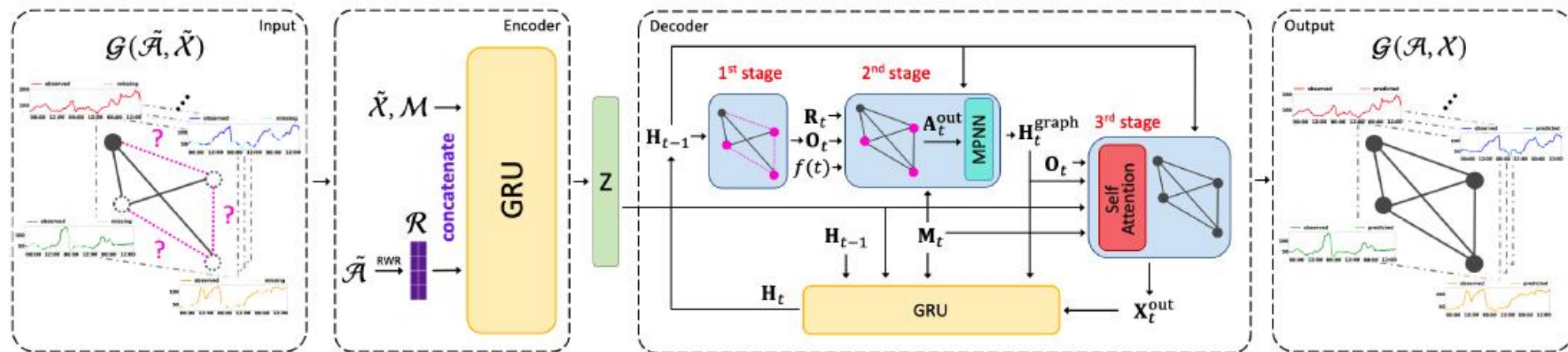


Figure 3: The model architecture of the proposed PoGeVon.

由于预测缺失节点特征和边缘的潜在互惠性，将NTS的输入作为一个**多任务学习问题**是很自然的，该问题由节点时间序列的**输入任务**和图结构的**链接预测任务**组成。

Multi-Task Learning Framework

首先从信息瓶颈角度分析NTS imputation建模为多任务学习的益处，我们把目标设定为 (1)。

$$\max[I(\tilde{\mathcal{A}}, \tilde{\mathcal{X}}; z) - \beta I(z; \tilde{\mathcal{G}}_{t:t+\Delta t})] \quad (1)$$

其中Z为隐状态，I为互信息，G则代表t~t+Δt的NTS数据滑动窗口的数据样本，β为拉格朗日乘子。

β I约束了每个数据样本中可以通过潜在表示传输的身份信息的数量。

而式中 $I(\tilde{\mathcal{A}}, \tilde{\mathcal{X}}; z)$ 则代表了VAE的重构任务，可分解为 (2)。

$$I(\tilde{\mathcal{A}}, \tilde{\mathcal{X}}; z) = I(\tilde{\mathcal{A}}; z) + I(\tilde{\mathcal{X}}; z) - I(\tilde{\mathcal{A}}; \tilde{\mathcal{X}}; z) \quad (2)$$

该互信息测量了多个不同随机变量之间的共享信息，特别的，当A和X相互独立是，我们有 (3)：

$$\begin{aligned} I(\tilde{\mathcal{A}}, \tilde{\mathcal{X}}; z) &= H(\tilde{\mathcal{A}}, \tilde{\mathcal{X}}) - H(\tilde{\mathcal{A}}, \tilde{\mathcal{X}}|z) \\ &= H(\tilde{\mathcal{A}}) + H(\tilde{\mathcal{X}}) - H(\tilde{\mathcal{A}}|z) - H(\tilde{\mathcal{X}}|z) = I(\tilde{\mathcal{A}}; z) + I(\tilde{\mathcal{X}}; z) \end{aligned} \quad (3)$$

METHODOLOGY

综上所述，我们的分析表明：

(1)当相邻节点的特征不相关时，我们可以独立地估算节点时间序列和图的邻接度(Eq. (3));

(2)然而，在实际应用中，节点特征和图结构往往是相互关联的(如图2)，在这种情况下，节点特征插入和图邻接插入之间可能存在相互增强的效应(Eq.(4))。

我们的分析还提供了新颖和关键的线索，可以指导设计用于学习多模态数据集的 **NTS** 编码器-解码器框架。

$$I(\tilde{\mathcal{A}}; \tilde{\mathcal{X}}; z) = I(\tilde{\mathcal{A}}; z) - I(\tilde{\mathcal{A}}; z | \tilde{\mathcal{X}}) = I(\tilde{\mathcal{X}}; z) - I(\tilde{\mathcal{X}}; z | \tilde{\mathcal{A}}) \quad (4)$$

对于编码器，Eq.(5)表明，潜表示(即编码器的输出)应同时编码图邻接信息和节点特征信息(即Eq.(5)的VAE部分)以及它们之间的相互作用(即Eq.(5)的CVAE部分)。

对于解码器，我们将提出一种三阶段预测方法，使得(输入的)图结构和(输入的)节点特征可以分别作为彼此的条件(即Eq.(5)的CVAE部分)。

$$2 \cdot I(\tilde{\mathcal{A}}, \tilde{\mathcal{X}}; z) = \underbrace{I(\tilde{\mathcal{A}}; z) + I(\tilde{\mathcal{X}}; z)}_{\text{VAE}} + \underbrace{I(\tilde{\mathcal{X}}; z | \tilde{\mathcal{A}}) + I(\tilde{\mathcal{A}}; z | \tilde{\mathcal{X}})}_{\text{Conditional VAE}} \quad (5)$$

Encoder

编码器旨在对NTS数据的结构信息和动态信息进行编码。现有的基于消息传递的gnn通常只从近邻捕获本地信息。

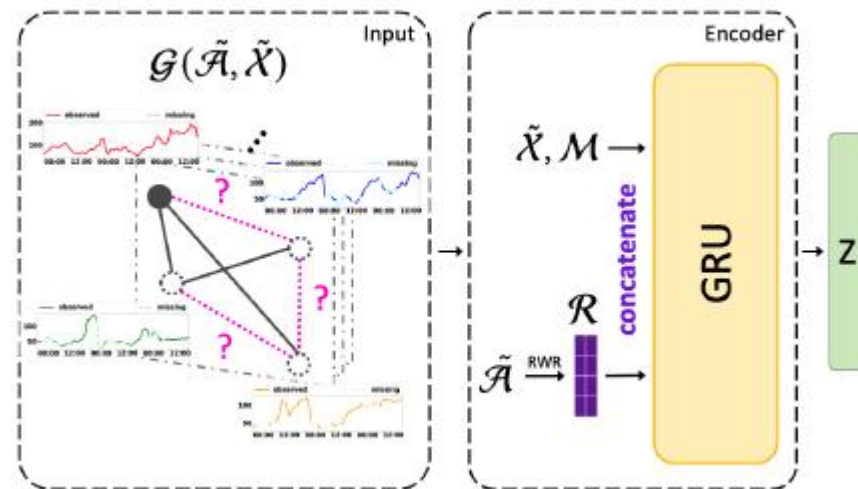
然而，节点之间的远距离信息在NTS数据中很重要，因为图是不断发展的，节点之间的交互可以在任何时间步长发生。

因此，为了获取远距离全局信息，方法使用随机行走重启(RWR)的位置嵌入，如（6）所示：

$$\mathbf{r}_{t,i} = (1 - c)\hat{A}_t \mathbf{r}_{t,i} + c\mathbf{e}_i \quad (6)$$

A为单位邻接矩阵，e为0-1向量，c为重启概率。

到达平稳分布后，我们将所有锚节点的位置分数r，作为最终的位置嵌入，其中N是节点的数目



METHODOLOGY

接下来，我们用下面的命题和定理证明了基于rw的位置嵌入的表达能Ⓕ力。

命题：在图学习中，随机重新启动游动(RWR)从近邻(局部)和远距离邻居(全局)中获取信息。

定理：在给定时间图G的情况下，基于rw的节点位置嵌入的TGN \diamond 具有比常规TGN \diamond \diamond 更强的表达能力节点表示学习： $D(\diamond(\diamond)\diamond(\diamond)) \geq D(\diamond(\diamond)\diamond(\diamond))$, D表达通过计算的节点对基于节点 $(G\diamond\diamond)$ 表示。

定理从基于消息传递的时间图网络(TGN)的角度总结了在时间图中基于rwr的位置嵌入的好处，TGN是设计用于处理时间图的通用gnn。它包含两个主要组成部分：

- 用于捕获每个节点动态的内存(通过rnn);
- 聚合和更新(通过gnn)来收集邻居的拓扑信息

METHODOLOGY

最后，为了捕获NTS数据中的动态信息，我们使用2层门控循环单元(GRU)作为编码器来建模 $(\chi | \sim X, M, R)$ ，其中 χ 是潜在表示， $R = \{R_1, \dots, R_{|\mathcal{N}|}\}$ 为节点位置嵌入张量。

对于每个 R_{i^*} ，不将所有节点都视为锚节点，通常只需锚节点的一小部分 $|\mathcal{A}_{i^*}| = |\mathcal{N}_{i^*}|$ 就足以在实践中区分节点。

在输入GRU之前， M 和位置嵌入 R 在每个时间步与输入 $\sim X$ 连接。

Decoder

我们将解码器设计为具有3阶段预测的GRU。

我们用 H_t 表示时间步长的节点嵌入矩阵 \diamond ， H 表示节点嵌入张量。

基于encoder的分析，我们对解码器中不同预测阶段的特征imputation \diamond 1和网络imputation \diamond 2之间的互补关系建模如下：

第一阶段特征预测：

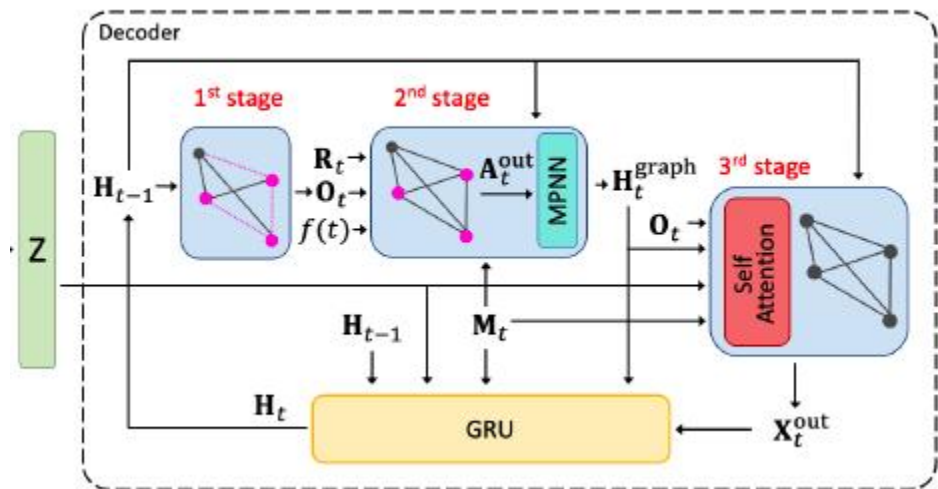
$$\text{线性层预测值: } \hat{Y}_{1,t} = \text{Linear}(H_{t-1}) \quad (7)$$

$$\text{输出值: } O_t = M_t \odot \tilde{X}_t + (1 - M_t) \odot \hat{Y}_{1,t} \quad (8)$$

第二阶段节点关系预测：

$$U_t = \text{Linear}(O_t \parallel M_t \parallel R_t \parallel H_{t-1}) \quad (9)$$

$$f(t) = \sqrt{\frac{1}{k}} [\cos(w_1 t), \sin(w_1 t), \dots, \cos(w_k t), \sin(w_k t)] \quad (10)$$



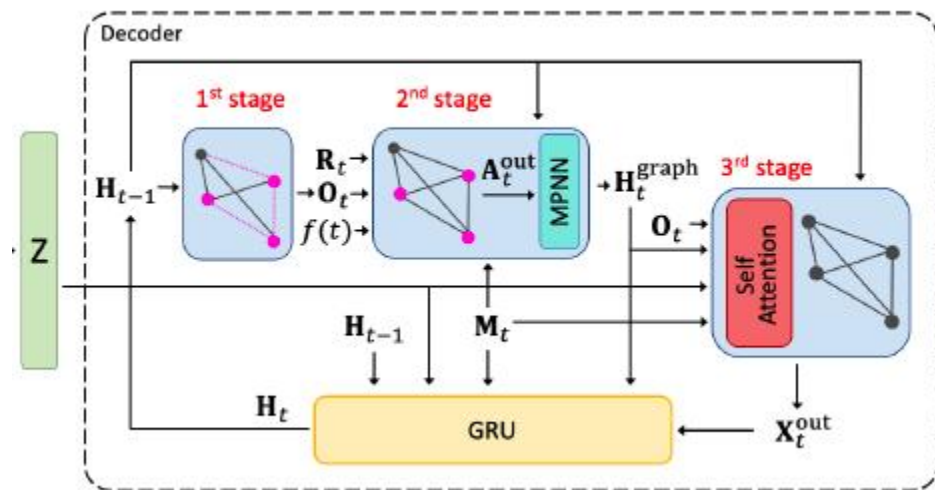
Decoder

直接使用之前时间步 H_{t-1} 的隐藏状态作为这些缺失节点的嵌入，因为在时间步 t 没有新的特征或图结构可用。在NTS中，观测值通常是通过不规则采样获得的，并且在现实世界问题的任何未来步骤中都可能出现对它们的imputation问题。

能够处理这种不确定性和预测未来时间步中看不见的图结构/时间序列数据是NTS imputation模型的两个关键特征。

因此，为了捕捉不同时间戳之间的动态，增强PoGeVon的表达能力，我们还基于Bochner定理，用可学习的傅立叶特征对时间信息进行编码

$$f(t) = \sqrt{\frac{1}{k}} [\cos(w_1 t), \sin(w_1 t), \dots, \cos(w_k t), \sin(w_k t)] \quad (10)$$



Decoder

然后，我们通过广播将节点嵌入与时间编码连接起来，作为两层多层感知器(MLP)的输入，以预测缺失边：

$$A_t^{\text{out}} = \text{MLP}(U_t \parallel H_{t-1} \parallel f(t)) \quad (11)$$

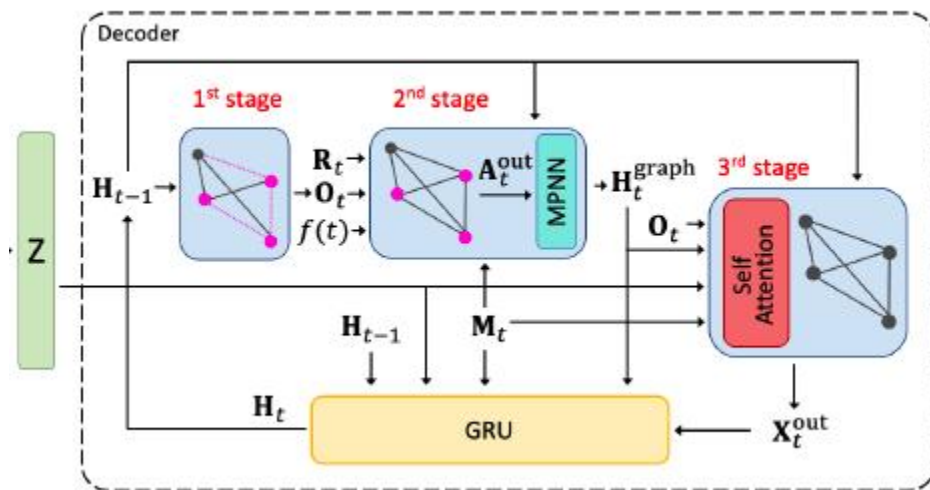
第三阶段特征预测：

在第三阶段预测中，文章利用结构信息 H_g 再次对节点特征时间序列进行细粒度的插值。

为了增强节点表示的语义，我们在第三阶段预测中应用了自关注层来捕获跨节点信息，这有助于编码 H_g 中未捕获的更丰富的节点交互信息。

$$H_t^{\text{out}} = \text{MLP}(\text{Attn}(Z \parallel H_{t-1} \parallel H_t^{\text{graph}} \parallel O_t \parallel M_t)) \quad (13)$$

$$\hat{Y}_{2,t} = \text{MLP}(H_t^{\text{out}} \parallel H_{t-1} \parallel H_t^{\text{graph}}) \quad (14)$$



Experiment

Experiment:

文章将提出的PoGeVon应用于网络时间序列imputation任务，并从以下几个方面对其进行评估:

- PoGeVon对网络时间序列imputation有多有效?
- 方法在多大程度上受益于模型的不同组成部分?

Dataset	# of entity	# of nodes	average # of edges	time length
COVID-19	50	50	1344.75	346
AQ36	36	36	341.57	8759
PeMS-BA	1632	64	675.45	25920
PeMS-LA	2383	64	1095.54	25920
PeMS-SD	674	64	1295.11	25920

表1: 所用数据集基本信息

Experiment

Experiment:

Models	COVID-19			AQ36		
	MAE	MSE	MRE	MAE	MSE	MRE
Mean	3.081 ± 0.000	10.707 ± 0.000	0.284 ± 0.000	62.299 ± 0.000	6525.709 ± 0.000	0.835 ± 0.000
MF	0.276 ± 0.026	0.165 ± 0.025	0.026 ± 0.002	39.582 ± 0.189	4545.596 ± 61.411	0.531 ± 0.002
MICE	0.077 ± 0.005	0.013 ± 0.002	0.007 ± 0.000	38.889 ± 0.268	4314.435 ± 20.617	0.521 ± 0.003
BRITS	0.386 ± 0.006	0.293 ± 0.009	0.036 ± 0.001	23.393 ± 0.802	1276.226 ± 102.916	0.314 ± 0.011
rGAIN	0.579 ± 0.069	0.571 ± 0.106	0.055 ± 0.006	25.032 ± 1.426	1358.134 ± 152.361	0.335 ± 0.019
SAITS	0.466 ± 0.010	0.366 ± 0.019	0.043 ± 0.001	51.097 ± 0.625	5026.475 ± 75.120	0.685 ± 0.008
TimesNet	0.028 ± 0.002	0.002 ± 0.000	0.003 ± 0.000	40.700 ± 0.278	3383.554 ± 49.499	0.545 ± 0.004
GRIN	0.319 ± 0.038	0.165 ± 0.040	0.029 ± 0.004	29.420 ± 0.231	2050.726 ± 56.028	0.394 ± 0.003
NET ³	0.547 ± 0.004	0.682 ± 0.006	0.051 ± 0.000	34.755 ± 0.497	2473.718 ± 37.461	0.466 ± 0.007
PoGeVON	0.007 ± 0.001	0.000 ± 0.000	0.001 ± 0.000	19.494 ± 1.101	1213.474 ± 125.529	0.261 ± 0.015

Models	PeMS-BA			PeMS-LA			PeMS-SD		
	MAE	MSE	MRE	MAE	MSE	MRE	MAE	MSE	MRE
Mean	192.047 ± 0.000	47504.159 ± 0.000	0.474 ± 0.000	216.681 ± 0.000	62664.657 ± 0.000	0.406 ± 0.000	208.192 ± 0.000	55780.002 ± 0.000	0.529 ± 0.000
MF	57.265 ± 1.148	8091.407 ± 185.123	0.141 ± 0.003	77.339 ± 0.699	15202.678 ± 156.348	0.145 ± 0.001	45.811 ± 0.318	6044.345 ± 72.976	0.117 ± 0.001
MICE	50.861 ± 0.765	6724.148 ± 109.829	0.126 ± 0.002	64.018 ± 1.015	10822.355 ± 405.410	0.120 ± 0.002	38.978 ± 1.036	4771.186 ± 92.335	0.100 ± 0.003
BRITS	30.274 ± 0.095	2942.411 ± 16.511	0.075 ± 0.000	36.921 ± 0.133	3681.595 ± 21.635	0.069 ± 0.000	21.232 ± 0.059	1563.234 ± 28.309	0.054 ± 0.000
rGAIN	38.862 ± 0.752	3422.914 ± 61.281	0.096 ± 0.002	49.611 ± 1.083	5533.964 ± 234.335	0.093 ± 0.002	33.212 ± 1.475	2341.466 ± 98.314	0.085 ± 0.004
SAITS	46.567 ± 0.530	5412.574 ± 161.132	0.115 ± 0.001	61.896 ± 0.892	10998.854 ± 204.345	0.116 ± 0.002	34.117 ± 0.886	4101.397 ± 152.141	0.087 ± 0.002
TimesNet	25.859 ± 0.115	1676.843 ± 16.144	0.064 ± 0.000	27.452 ± 0.114	2058.227 ± 6.213	0.052 ± 0.000	21.583 ± 0.085	1284.300 ± 21.839	0.055 ± 0.000
GRIN	30.057 ± 1.073	1922.072 ± 74.327	0.074 ± 0.003	47.835 ± 2.059	4561.512 ± 298.533	0.090 ± 0.004	41.001 ± 1.543	3000.012 ± 201.018	0.105 ± 0.004
NET ³	35.671 ± 0.111	2735.574 ± 6.138	0.009 ± 0.000	37.652 ± 0.113	3416.784 ± 6.765	0.071 ± 0.000	34.111 ± 0.184	2487.581 ± 9.798	0.087 ± 0.000
PoGeVON	22.194 ± 0.046	1248.681 ± 4.297	0.055 ± 0.000	23.905 ± 0.245	1714.962 ± 31.035	0.045 ± 0.000	18.990 ± 0.112	951.559 ± 8.264	0.048 ± 0.000

表2: 基线在5种数据集上与不同方法的性能对比

Experiment

Experiment:

Ablation study & sensitivity analyze

Models	MAE	MSE	MRE
PoGeVON	19.49 ± 1.10	1213.47 ± 125.53	0.26 ± 0.02
change RWR to SPD	21.98 ± 1.55	1309.55 ± 199.24	0.33 ± 0.02
change RWR to RWPE Embeddings	23.75 ± 0.85	1597.67 ± 210.77	0.32 ± 0.01
change RWR to PGNN Embeddings	24.46 ± 2.59	1625.19 ± 393.25	0.33 ± 0.04
w/o link prediction in 2 nd stage	28.71 ± 3.38	2130.46 ± 417.45	0.38 ± 0.05
w/o self-attention in 3 rd stage	23.40 ± 1.00	1576.06 ± 194.45	0.31 ± 0.01

