计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
62(2): 346−363, 2025

融合不确定性建模的时空交通数据插补方法

刘 乐[1,2] 郭晟楠[1,2] 靳希源[1,2] 赵苗苗[1] 陈 冉[1] 林友芳[1,2] 万怀宇[1,2]

[1](北京交通大学计算机与信息技术学院 北京 100044)

[2](交通数据分析与挖掘北京市重点实验室(北京交通大学) 北京 100044)

(liulecs@bjtu.edu.cn)

Spatial-Temporal Traffic Data Imputation Method with Uncertainty Modeling

Liu Le[1,2], Guo Shengnan[1,2], Jin Xiyuan[1,2], Zhao Miaomiao[1], Chen Ran[1], Lin Youfang[1,2], and Wan Huaiyu[1,2]

[1] (*School of Computer and Information Technology*, *Beijing Jiaotong University*, *Beijing* 100044)

[2] (*Beijing Key Laboratory of Traffic Data Analysis and Mining* (*Beijing Jiaotong University*), *Beijing* 100044)

Abstract Traffic data missing is one of the unavoidable problems in intelligent transportation systems. Completing and quantifying the uncertainty of missing values can improve the performance and reliability of traffic data mining tasks in intelligent transportation systems. However, most existing traffic data imputation models mainly focus on point estimation without quantifying the uncertainty, so they cannot meet the need for traffic data reliability in the transportation field. Besides, these methods only focus on modeling spatial-temporal correlation of traffic data, failing to consider the impact of missing values on spatial-temporal correlation. In addition, the uncertainty of traffic data is affected by time, spatial location, and the state of the data, but existing methods cannot comprehensively consider these factors. To address these challenges, we propose a spatial-temporal uncertainty guided traffic data imputation network (STUIN), which simultaneously realizes the imputation of spatial-temporal traffic data and the uncertainty quantification of the imputation results by self-supervised training. Specifically, we innovatively model the hidden states of the neural network as random variables subject to Gaussian distributions, use the variances of Gaussian distributions to model the uncertainty of the hidden states, and introduce a variance-based attention mechanism to characterize the effect of uncertainty on modeling spatio-temporal correlations. In addition, we design a novel spatial-temporal uncertainty initialization module, which incorporates the influence of time, space and missing values when initializing the means and variances of the Gaussian distributions. Experiments on two traffic flow datasets show that STUIN achieves state-of-the-art performance on both the data imputation and uncertainty quantification tasks.

Key words traffic data imputation; uncertainty quantification; variance-based attention mechanism; spatial-temporal data mining; spatial-temporal graph data

摘 要 交通数据缺失是智能交通系统无法避免的问题之一,对缺失值进行补全和不确定性量化能提高智能交通系统中交通数据挖掘相关任务的精度和可靠性.然而,目前大多数交通数据插补模型都只能针对缺失值给出点估计,无法量化不确定性,难以满足交通领域对数据可靠性的要求.而且,现有方法将重点放在了建模交通数据的时空相关性上,却未能在捕获时空相关性的过程中考虑缺失值的影响.此外,交通数据的不确定性同时受到时间、空间位置以及数据自身状态的影响,但是现有方法无法全面考虑这些因素的影响.为了解决这些问题,提出了一种时空不确定性指导的交通数据插补模型(spatial-temporal

收稿日期: 2023-06-05; **修回日期**: 2024-01-11

基金项目: 国家自然科学基金青年基金项目(62202043)

This work was supported by the National Natural Science Foundation of China for Young Scientists (62202043).

通信作者: 郭晟楠(guoshn@bjtu.edu.cn)

uncertainty guided traffic data imputation network, STUIN)，以自监督训练的方式实现了时空交通数据的插补和对插补结果的不确定性量化. 具体来说，创新地将神经网络的隐状态建模成服从高斯分布的随机变量，借助方差建模隐状态的不确定性，利用基于方差的注意力机制描述不确定性对时空相关性建模的影响；此外，设计了一个新颖的时空不确定性初始化模块，在初始化均值和方差时同时考虑了时间、空间和数据缺失状况多种因素的影响. 在 2 个交通流量数据集上的实验结果表明 STUIN 在数据插补和不确定性量化上都达到了最先进的性能.

关键词 交通数据插补；不确定性量化；基于方差的注意力机制；时空数据挖掘；时空图数据

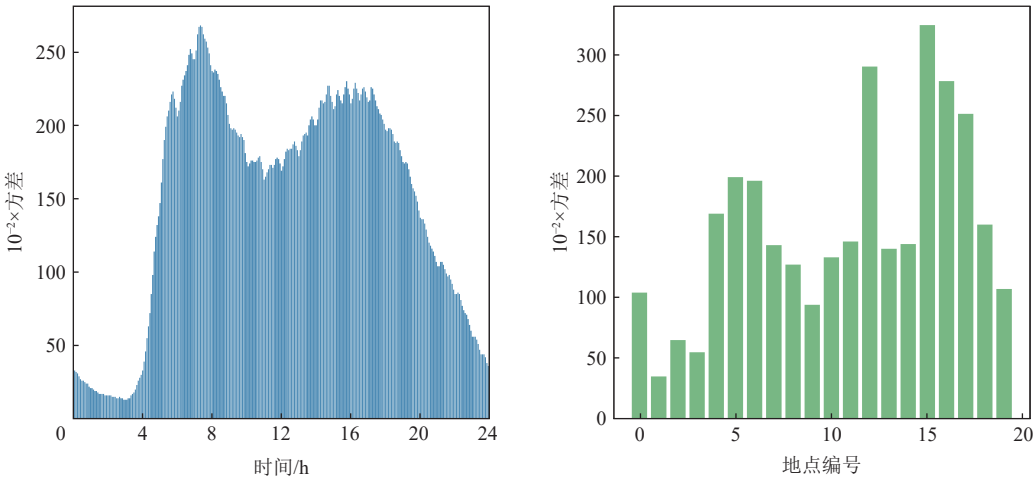
中图法分类号 TP183

DOI: 10.7544/issn1000-1239.202330455 **CSTR:** 32373.14.issn1000-1239.202330455

随着信息技术的发展，人们可以收集大量的交通数据，并将这些数据应用于车流量预测、出行需求预测等重要任务. 这些任务是智能交通系统的核心组件，极大地便利了人们的生活. 但是由于设备故障、通信不稳定等原因，实际收集到的交通数据通常是不完整的. 例如，据统计北京每日交通量的平均数据缺失率约为 10%[1]. 在实际应用场景中，数据的准确性和完整性对交通服务的质量有极大影响. 因此在将交通数据应用于各种下游智能化应用任务之前，必须先对其中的缺失值进行补全.

尽管已经有一些学者对交通数据插补进行了深入的研究，但这些方法通常都只能针对缺失值给出点估计，无法量化插补结果的不确定性[2]，即只给出了插补值的期望，而没有量化插补值的波动区间. 插补值的不确定性一定程度上体现了插补结果的可信赖程度. 若插补值的不确定性过大，插补结果则具有较低的可信度，此时若直接将其应用于下游任务，不但不能有效提升输入到下游任务中的数据质量，甚至可能会导致下游任务给出偏差较大的输出，进而导致严重的后果. 因此，在对交通数据进行插补的同时量化插补结果的不确定性，为插补结果给出置信区间，是一个重要且具有实际意义的研究问题. 准确的交通数据插补和不确定性量化，可以提高插补结果的可靠性，更好地支撑下游基于完整交通数据的服务应用. 但是，实现该任务具有 3 点挑战：

1)在交通数据中，不确定性会同时受到时间、空间和数据本身这 3 个因素的影响. 具体来说，不同时间段、不同空间位置的交通数据具有固有的模式特点，即交通数据具有时空异质性[3]. 例如本文在 PEMS04[4] 数据集上分别计算了时间维度和空间维度的方差，将计算结果绘制在了图 1 中. 在时间维度上，本文将 1 天平均划分为 288 个时间片，计算每个时间片内所有数据的方差，从图 1 可以看出，通勤高峰期的方差显著大于其他时间段的方差，因为此时更易发生交通拥堵和事故. 在空间维度上，本文计算每个传感器记录的所有数据的方差，任意挑选 20 个绘制在图 1



(a) 时间维度的方差 (b) 空间维度的方差

Fig. 1 Variance statistics of temporal and spatial dimensions

图 1 时间和空间维度的方差统计

中, 可以看出不同位置的方差具有显著的差异, 因此交通数据在空间维度上具有显著的异质性. 此外, 数据本身的状态, 即是否缺失, 也会影响该处值的不确定性, 若某处发生了缺失, 它的不确定性必然会更加明显. 因此, 在交通数据插补任务中量化不确定性需要考虑时间、空间、数据本身状态等多种因素的影响, 如何将这些因素综合起来是一个具有挑战性的问题.

2)对交通数据进行补全的核心在于挖掘和利用带缺失的交通数据中蕴含的时空相关性, 对于这一点, 目前尚没有完善的解决方案. 单纯地捕获时空相关性是时空预测[5]任务的研究重点, 并且该任务已经有了很多成熟的解决方案[3-4,6]. 尽管如此, 时空预测模型并不能直接用于捕获带缺失数据的时空相关性. 因为时空预测任务假定所有的数据都是有效的、可信的, 因此所有的数据点都可以在时间维度和空间维度进行信息的有效传播. 但时空交通数据插补任务中的输入数据中存在大量缺失值, 这些缺失值不是有效数据, 因此在建模时空相关性和进行时间和空间维度的消息传播时, 需要特殊处理缺失位置的数据. 例如若某个节点在某时刻发生了缺失, 那么与之相邻的节点就不应该直接从该节点聚合信息, 如何解决该问题值得进一步思考.

3)现实情况中缺失值是不存在真实标签的, 因此依赖缺失数据真实标签的有监督训练方式在实际中是无法落地应用的. 目前很多插补模型的解决方案是以自编码器的模式训练, 即仅让模型重构非缺失值, 但这种方式很难保证模型在缺失值上具有良好的泛化能力. 因此, 需要寻找一种新的模型训练范式, 在不依赖缺失值真实标签的前提下, 获得泛化能力强的插补模型.

针对以上问题, 本文提出了一种时空不确定性指导的交通数据插补模型(spatial-temporal uncertainty guided traffic data imputation network, STUIN), 该模型采用自监督的训练方式, 通过 Mask 非缺失值生成训练标签; 并将神经网络的隐状态建模成服从高斯分布的随机变量来量化不确定性, 削弱无效信息传播, 准确建模时空相关性, 并且在建模不确定性时同时考虑了时间、空间和数据缺失多种因素的影响. 本文的主要贡献有 3 点:

1)设计了一个新颖的时空不确定性初始化模块, 用于初始化被建模为高斯分布的隐状态的均值和方差, 该模块同时考虑了时间异质性、空间异质性和数据缺失状况.

2)为了准确建模带缺失的交通数据的时空相关性, 本文在时间和空间维度上设计了一种新颖的基于方差的注意力机制, 显式利用隐状态的不确定性指导消息在时间和空间维度的传播.

3)在 2 个交通流量数据集上构造了随机缺失、非随机缺失和全盲缺失 3 种缺失模式, 并设置了 4 个缺失率进行实验, 实验结果表明 STUIN 在数据插补和不确定性量化上都达到了最优的性能.

1 相关工作

1.1 插 补

缺失值插补是一个被广泛研究的问题, 可以将现有方法分为 3 种: 传统方法、机器学习方法和深度学习方法. 传统方法一般使用零值或其他统计特征值来替代缺失值, 效果往往不佳.

在统计机器学习领域, Lebel 等人[7]将自回归移动平均模型(autoregressive integrated moving average model, ARIMA) 拓展到插补领域. Hudak 等人[8]使用 K 近邻(K-nearest neighbors, KNN)方法进行缺失值插补. Chen 等人[9]提出了一种低秩自回归张量补全(low-rank autoregressive tensor completion, LATC)框架, 将时间变化作为正则化项引入到 3 维张量的补全中, 并设计了一种交替优化方案解决优化问题.

在深度学习领域, Mattei 等人[10]将重要性加权自编码器拓展到了带缺失的数据集上, 实现缺失数据补全. Che 等人[11]提出 GRU-D 模型处理带缺失值的时间序列, 利用缺失模式对门控循环单元(gated recurrent unit, GRU)的隐状态进行衰减, 削弱缺失值的影响, 但这种机制很难拓展到时空数据. Cao 等人[12]提出了 BRITS 模型, BRITS 改进了 GRU-D 的衰减机制, 使用双向循环神经网络对 GRU-D 进行增强. Shukla 等人[13]提出了一种处理不规则采样时间序列的方法, 采用了编码器-解码器结构, 编码器通过一组参考时间片来表示整个时间序列, 解码器从参考时间片中恢复时间序列. 文献 [10-13] 方法针对的都是时间序列, 在时空数据领域, Wu 等人[14]通过采样获得子图和相应的邻接矩阵, 然后使用扩散图卷积[15]重构缺失数据. Xu 等人[16]通过相关系数重构邻接矩阵, 使用 GraphSAGE[17]模型学习空间相关性, 并使用生成对抗网络(generative adversarial network, GAN)来插补数据. Ye 等人[18]提出了一种图注意力卷积网络(graph attention convolutional network, GACN), GACN 采用了编码器-解码器结构, 使用图注意力网络(graph attention

network, GAT)学习空间相关性和使用时间卷积学习时间相关性. 但这些时空数据插补方法都主要关注捕获时空相关性, 忽略了缺失值的影响.

1.2 不确定性量化

现有的不确定性量化方法可以分为2类: 基于贝叶斯的和基于非贝叶斯的. 基于贝叶斯的方法为模型参数指定先验分布, 然后推断网络参数的后验分布来量化认知不确定性, 其实现难点在于后验分布难以直接计算, 只能使用近似方法. 如变分推断[19-21]和马尔可夫链蒙特卡罗方法[22-23], 但它们通常都比较耗时. Gal等人[24-25]提出了蒙特卡洛丢弃法(Monte Carlo dropout, MC Dropout), 将变分分布指定为伯努利分布, MC Dropout不需要改变现有的模型结构, 只需在每个隐藏层之后应用随机Dropout, 模型输出可以近似地视为后验分布生成的随机样本.

相比于基于贝叶斯的方法, 基于非贝叶斯的方法更加灵活. 如平均间隔分数(mean interval score, MIS)回归[26]和分位数回归[27]在损失函数中融入了不确定性度量. 样条分位数(spline quantile, SQ)回归[28]对每个点估计使用若干参数来构造分位数函数, 通过最小化连续概率排位分数(continuous ranked probability score, CRPS)使分位数函数拟合训练集, 使用分位数函数量化不确定性. Salinas等人[29]提出了DeepAR模型对时间序列的概率分布进行估计. Zhou等人[30]提出了时空不确定性感知预测网络(spatio temporal uncertainty-aware prediction network, STUaNet), 设计了一种不确定性学习机制, 可以同时估计内部数据质量和量化外部不确定性.

虽然不确定性量化早已被广泛研究, 但现有方法大多只适用于完整数据集, 在缺失数据集中量化不确定性必须考虑缺失值的影响, 对时空数据来说, 时空相关性的存在则使该问题变得更加复杂.

2 相关概念和问题定义

本节将对本文涉及的概念和要解决的问题进行形式化定义.

2.1 基本定义

定义 1. 交通路网. 交通路网为一个有向或无向图$G=(V,E)$. 其中V为$|V|=N$个节点的集合, 每个节点代表交通路网中的一个观测点; E为节点间边的集合; $\boldsymbol{A}\in\mathbb{R}^{N\times N}$为$G$的邻接矩阵, 若$e_{ij}\in E$则$A_{ij}=1$.

定义 2. 交通观测向量. 交通路网在第t个时间片的观测向量为$\boldsymbol{x}_{:,t}=(x_{1,t},\cdots,x_{N,t})^{\mathrm{T}}\in\mathbb{R}^N$, 其中$x_{v,t}$表示节点$v$在第$t$个时间片的观测值, 交通观测向量带有缺失值.

定义 3. 掩码向量. 掩码向量$\boldsymbol{m}\in\{0,1\}^N$用于指示第$t$个时间片的交通观测向量的缺失情况, 若$x_{v,t}$缺失则$m_{v,t}=0$, 反之$m_{v,t}=1$.

2.2 问题定义

交通数据插补和不确定性量化. 给定连续T个时间片的交通观测矩阵$\boldsymbol{X}=(\boldsymbol{x}_{:,1},\cdots,\boldsymbol{x}_{:,T})\in\mathbb{R}^{N\times T}$, 用掩码矩阵$\boldsymbol{M}=(\boldsymbol{m}_{:,1},\cdots,\boldsymbol{m}_{:,T})\in\{0,1\}^{N\times T}$指示其缺失情况, 本文的目标是对$\boldsymbol{X}$中的缺失值进行补全, 同时输出补全结果的不确定性: $\boldsymbol{Y},\boldsymbol{\Sigma}=f(\boldsymbol{X},\boldsymbol{M})$, 其中$\boldsymbol{Y}$表示插补结果, $\boldsymbol{\Sigma}$为不确定性度量.

3 时空不确定性指导的交通数据插补模型

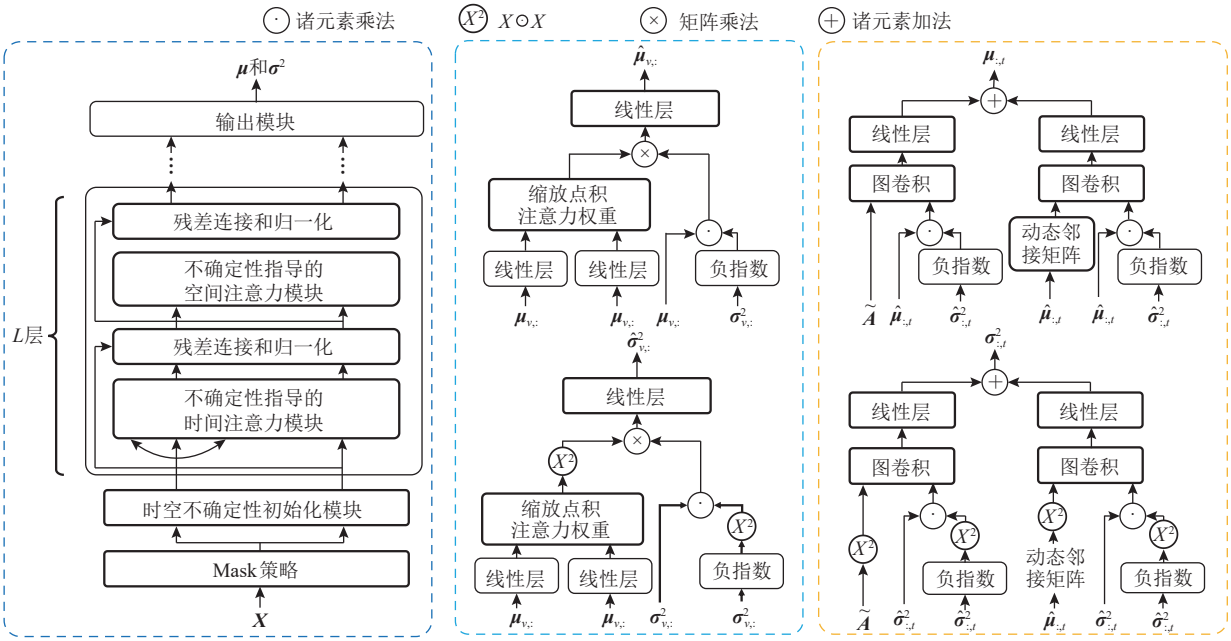
3.1 模型整体结构

STUIN的模型结构如图2所示. 总体来说, 模型采用自监督学习的方式进行训练. 首先通过Mask策略删除$\boldsymbol{X}$中的一部分非缺失值得到$\tilde{\boldsymbol{X}}$和相应的掩码矩阵$\tilde{\boldsymbol{M}}$, 利用这部分非缺失值的真实值指导模型训练. 之后通过时空不确定性初始化模块得到隐状态的均值$\boldsymbol{\mu}^{(0)}\in\mathbb{R}^{N\times T\times d}$和方差$\boldsymbol{\sigma}^{(0)^2}\in\mathbb{R}^{N\times T\times d}$. 然后将$\boldsymbol{\mu}^{(0)}$和$\boldsymbol{\sigma}^{(0)^2}$输入不确定性指导的时空注意力模块, 该模块由多个相同的层堆叠而成, 每层包含1个不确定性指导的时间注意力模块和1个不确定性指导的空间注意力模块. 这2个模块在学习时空相关性时引入了基于方差的注意力机制, 使用基于方差的权重对隐状态中的均值和方差加权, 方差越大则权重越小, 缺失值所代表的噪声信号往往不确定性较高, 具有较大的方差. 因此, 可以通过这种方式削弱缺失值带来的无效信息传播, 实现利用不确定性指导模型学习时空相关性. 最后通过输出模块得到模型输出$\boldsymbol{\mu}$和$\boldsymbol{\sigma}^2$, 将$\boldsymbol{\mu}$作为插补结果, 使用$\boldsymbol{\sigma}^2$量化不确定性.

3.2 时空不确定性初始化模块

该模块根据输入数据得到隐状态的均值和方差, 在具体实现时, 本文通过学习不同的嵌入矩阵实现在量化不确定性时融合不同因素的影响.

首先, 每个交通节点都与一些静态特征相关联, 例如局部拓扑结构、所在道路类型等, 这些特征能增强模型对空间相关性的捕获能力. 然而, 在现实条件下, 很难获取每个节点的所有静态特征. 因此, 本文定义空间嵌入矩阵$\boldsymbol{SE}\in\mathbb{R}^{N\times 1\times d}$, 为每个节点赋予一个$d$维的嵌入向量, 让其参与均值和方差的初始化, 显式地建模空间异质性. 类似地, 不同时间片也具有不



(a) STUIN模型整体结构 (b) 不确定性指导的时间注意力模块 (c) 不确定性指导的空间注意力模块

Fig. 2 Structure of STUIN model

图 2 STUIN 模型结构

同的静态特征, 可以通过时间片嵌入来显式地建模时间异质性, 然而, 对于插补任务而言, 某一时刻的值通常仅与相邻时刻具有较大的关联, 与较远时刻的值几乎无关. 因此, 在时间维度上相对性更为重要, 没有必要建模绝对时间异质性. 由于每个样本包含 T 个时间片, 本文通过时间嵌入矩阵 $\boldsymbol{TE}\in\mathbb{R}^{1\times T\times d}$ 定义了 T 个时间片嵌入向量, 用于初始化均值和方差.

其次, 自注意力机制的置换不变性会导致模型无法感知时间序列的位置关系, 因此, 需要显式地附加位置编码. 本文通过 $\boldsymbol{TPE}\in\mathbb{R}^{1\times T\times d}$ 对均值附加固定位置编码[31], 其中位置 t 的编码由式(1)确定:

$$\begin{aligned} TPE_{t,2i} &= \sin(t/10\,000^{2i/d}),\\ TPE_{t,2i+1} &= \cos(t/10\,000^{2i/d}). \end{aligned} \tag{1}$$

最后, 在模型的定义中, 均值对应于从数据中提取的特征. 因此, 在初始化均值时需要附加上输入数据的信息, 本文采用前馈神经网络来实现该操作:

$$g(\tilde{\boldsymbol{X}}) = \tilde{\boldsymbol{X}}\boldsymbol{W}_{\mathrm{E}} + \boldsymbol{b}_{\mathrm{E}}, \tag{2}$$

其中 $\boldsymbol{W}_{\mathrm{E}}\in\mathbb{R}^{1\times d}$ 和 $\boldsymbol{b}_{\mathrm{E}}\in\mathbb{R}^{d}$ 是可学习参数. 方差用于建模不确定性, 而不确定性与数据是否缺失相关联. 因此, 本文定义了 2 个 d 维嵌入向量表示缺失和非缺失 2 种状态, 用于初始化方差, 并定义方差嵌入函数 $VE(\cdot)$, 根据数据的缺失状况获取相应的嵌入向量.

综上所述, 该模块的计算公式可以概括为

$$\begin{aligned} \boldsymbol{\mu}^{(0)} &= \boldsymbol{TE}+\boldsymbol{SE}+\boldsymbol{TPE}+g(\tilde{\boldsymbol{X}}),\\ \boldsymbol{\sigma}^{(0)^2} &= softplus(\boldsymbol{TE}+\boldsymbol{SE}+VE(\tilde{\boldsymbol{M}})), \end{aligned} \tag{3}$$

其中 $softplus(\cdot)$ 是激活函数, 定义为

$$softplus(x) = \ln(1+\exp(x)), \tag{4}$$

可以视为 $relu(\cdot)$ 激活函数的平滑版本.

3.3 不确定性指导的时空注意力模块

该模块由多个相同的层堆叠而成, 每一层由不确定性指导的时间注意力模块和不确定性指导的空间注意力模块构成, 并在它们之间使用残差连接和归一化解决梯度消失问题, 加快模型收敛速度.

3.3.1 不确定性指导的时间注意力模块

该模块使用多头自注意力机制学习时间相关性, 并在其中结合基于方差的注意力机制. 为方便阅读和理解, 本节仅介绍该模块在单个注意力头下的计算方式, 多个注意力头下的计算方式在阅读本节后容易得出, 并且本节省略表示层数的上标. 设节点 v 在某一层的隐状态为均值 $\boldsymbol{\mu}_{v,:}\in\mathbb{R}^{T\times d}$ 和方差 $\boldsymbol{\sigma}^2_{v,:}\in\mathbb{R}^{T\times d}$, 首先计算基于方差的注意力权重:

$$\boldsymbol{\beta}_{v,:} = \exp(-\tau\boldsymbol{\sigma}^2_{v,:}), \tag{5}$$

其中 $\tau>0$ 为超参数. 显然, 方差越大, 注意力权重越小, 后文将使用 $\boldsymbol{\beta}_{v,:}$ 通过诸元素乘积的形式对隐状态加权. 然后根据 $\boldsymbol{\mu}_{v,:}$ 计算查询矩阵 $\boldsymbol{Q}$ 和键矩阵 $\boldsymbol{K}$:

$$\begin{aligned} \boldsymbol{Q} &= \boldsymbol{\mu}_{v,:}\boldsymbol{W}_{\boldsymbol{Q}}\in\mathbb{R}^{T\times d_k},\\ \boldsymbol{K} &= \boldsymbol{\mu}_{v,:}\boldsymbol{W}_{\boldsymbol{K}}\in\mathbb{R}^{T\times d_k}, \end{aligned} \tag{6}$$

其中 $\boldsymbol{W}_{\boldsymbol{Q}}\in\mathbb{R}^{d\times d_k}$ 和 $\boldsymbol{W}_{\boldsymbol{K}}\in\mathbb{R}^{d\times d_k}$ 为可学习参数. 根据 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 计算自注意力权重 $\boldsymbol{E}$:

$$\boldsymbol{E}=softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right). \tag{7}$$

若 2 个随机变量相互独立且服从高斯分布，它们的加权和也服从高斯分布. 而自注意力机制和基于方差的注意力机制均遵循这种计算方式，因此若假设节点 v 在不同时间片的隐状态相互独立，则计算后的隐状态仍服从高斯分布，具体计算公式为：

$$\begin{aligned}\boldsymbol{z}_{\mu}&=\boldsymbol{E}(\boldsymbol{\beta}_{v,:}\odot\boldsymbol{\mu}_{v,:}),\\ \boldsymbol{z}_{\sigma}&=(\boldsymbol{E}\odot\boldsymbol{E})(\boldsymbol{\beta}_{v,:}\odot\boldsymbol{\beta}_{v,:}\odot\boldsymbol{\sigma}^2_{v,:}).\end{aligned} \tag{8}$$

即先利用 $\boldsymbol{\beta}_{v,:}$ 对均值和方差加权，再进行自注意力机制的计算. 之后还需对 $\boldsymbol{z}_{\mu}$ 和 $\boldsymbol{z}_{\sigma}$ 施加线性映射和非线性激活函数，但该操作将使后续的分布难以求解，为了简便，本节直接对 $\boldsymbol{z}_{\mu}$ 和 $\boldsymbol{z}_{\sigma}$ 施加这些操作，则此模块的输出为

$$\begin{aligned}\hat{\boldsymbol{\mu}}_{v,:}&=relu(\boldsymbol{z}_{\mu}\boldsymbol{W}^{\mu}_{\mathrm{T}}),\\ \hat{\boldsymbol{\sigma}}^2_{v,:}&=softplus(\boldsymbol{z}_{\sigma}\boldsymbol{W}^{\sigma}_{\mathrm{T}}),\end{aligned} \tag{9}$$

其中 $\boldsymbol{W}^{\mu}_{\mathrm{T}}\in\mathbb{R}^{d\times d}$，$\boldsymbol{W}^{\sigma}_{\mathrm{T}}\in\mathbb{R}^{d\times d}$ 是可学习参数.

3.3.2　不确定性指导的空间注意力模块

该模块使用图卷积学习空间相关性，同时结合基于方差的注意力机制. 为了便于阅读，本节同样省略表示层数的上标. 设均值 $\hat{\boldsymbol{\mu}}_{:,t}\in\mathbb{R}^{N\times d}$ 和方差 $\hat{\boldsymbol{\sigma}}^2_{:,t}\in\mathbb{R}^{N\times d}$ 是某一层第 t 个时间片所有节点的隐状态，首先计算基于方差的注意力权重：

$$\boldsymbol{\alpha}_{:,t}=\exp(-\nu\hat{\boldsymbol{\sigma}}^2_{:,t}), \tag{10}$$

其中 $\nu>0$ 为超参数. 与 3.3.1 节的处理方式相同，本文假设第 t 个时间片所有节点的隐状态是相互独立的，则图卷积后的隐状态计算公式为

$$\begin{aligned}\boldsymbol{c}_{\mu}&=\tilde{\boldsymbol{A}}(\boldsymbol{\alpha}_{:,t}\odot\hat{\boldsymbol{\mu}}_{:,t}),\\ \boldsymbol{c}_{\sigma}&=(\tilde{\boldsymbol{A}}\odot\tilde{\boldsymbol{A}})(\boldsymbol{\alpha}_{:,t}\odot\boldsymbol{\alpha}_{:,t}\odot\hat{\boldsymbol{\sigma}}^2_{:,t}),\end{aligned} \tag{11}$$

其中 $\tilde{\boldsymbol{A}}$ 为归一化后的邻接矩阵.

式(11)只使用了原始邻接矩阵进行图卷积，原始的邻接矩阵通常不能完全反映节点之间的相关性，甚至与真实的节点相关性相违背. 此外，原始邻接矩阵是静态的，在实际情况中，空间相关性可能会随时间动态变化. 因此本文引入动态邻接矩阵：

$$\boldsymbol{S}_t=softmax\left(\frac{\hat{\boldsymbol{\mu}}_{:,t}\hat{\boldsymbol{\mu}}^{\mathrm{T}}_{:,t}}{\sqrt{d}}\right), \tag{12}$$

其中 S_{ij} 表示节点 i 和节点 j 的相关性，值越大则相关性越大. 之后使用动态邻接矩阵进行图卷积：

$$\begin{aligned}\boldsymbol{c}^{\mathrm{a}}_{\mu}&=\boldsymbol{S}_t(\boldsymbol{\alpha}_{:,t}\odot\hat{\boldsymbol{\mu}}_{:,t}),\\ \boldsymbol{c}^{\mathrm{a}}_{\sigma}&=(\boldsymbol{S}_t\odot\boldsymbol{S}_t)(\boldsymbol{\alpha}_{:,t}\odot\boldsymbol{\alpha}_{:,t}\odot\hat{\boldsymbol{\sigma}}^2_{:,t}),\end{aligned} \tag{13}$$

并对 2 次图卷积的结果施加线性映射和非线性激活函数：

$$\begin{aligned}\boldsymbol{\mu}_{:,t}&=relu(\boldsymbol{c}_{\mu}\boldsymbol{W}^{\mu}_{\mathrm{S}})+relu(\boldsymbol{c}^{\mathrm{a}}_{\mu}\boldsymbol{W}^{\mu}_{\mathrm{A}}),\\ \boldsymbol{\sigma}^2_{:,t}&=softplus(\boldsymbol{c}_{\sigma}\boldsymbol{W}^{\sigma}_{\mathrm{S}})+softplus(\boldsymbol{c}^{\mathrm{a}}_{\sigma}\boldsymbol{W}^{\sigma}_{\mathrm{A}}),\end{aligned} \tag{14}$$

其中 $\boldsymbol{W}^{\mu}_{\mathrm{S}}$，$\boldsymbol{W}^{\mu}_{\mathrm{A}}$，$\boldsymbol{W}^{\sigma}_{\mathrm{S}}$，$\boldsymbol{W}^{\sigma}_{\mathrm{A}}$ 为可学习参数.

3.3.3　残差连接与归一化

由于本文将神经网络的隐状态定义成服从高斯分布的随机变量. 因此，需要对残差连接和归一化进行拓展. 设 $\boldsymbol{\mu}$ 和 $\boldsymbol{\sigma}^2$ 为某一个模块的输出，$\boldsymbol{\mu}'$ 和 $\boldsymbol{\sigma}'^2$ 为该模块的输入. 则采用层归一化对均值做归一化，不对方差做归一化，因为对方差进行归一化会带来 2 个问题：1)方差在模型中代表可信度，归一化会破坏它的数量级；2)归一化后会导致方差小于 0，后续模块无法直接使用归一化后的方差，如何处理负方差也是一个难题.

$$\begin{aligned}\boldsymbol{\mu}&\leftarrow normalize(\boldsymbol{\mu}),\\ \boldsymbol{\sigma}^2&\leftarrow\boldsymbol{\sigma}^2.\end{aligned} \tag{15}$$

实验中也发现对方差进行归一化会导致模型性能下降，而完全不进行归一化会出现梯度爆炸问题. 之后对均值和方差都施加残差连接操作：

$$\begin{aligned}\boldsymbol{\mu}&\leftarrow\boldsymbol{\mu}+\boldsymbol{\mu}',\\ \boldsymbol{\sigma}^2&\leftarrow\boldsymbol{\sigma}^2+\boldsymbol{\sigma}'^2.\end{aligned} \tag{16}$$

3.4　自监督训练策略

自监督训练策略包括 Mask 策略和损失函数，Mask 策略用于生成训练标签，损失函数用于衡量模型对这些标签的重构效果.

3.4.1　Mask 策略

本文通过 Mask 原始数据中的非缺失值构造训练标签，为了使构造的缺失值和原始数据具有相同的缺失模式，本文使用的具体策略如下[32]：设当前样本为 A，从数据集中随机选择一个样本 B，若样本 B 在某个位置缺失，则将样本 A 对应位置也设为缺失. 由于同一数据集中的样本通常具有相同的缺失模式，因此构造的训练标签能够与数据集的原始缺失模式保持一致，其示意图如图 3 所示.

值得注意的是，使用上述 Mask 策略时，对一个样本 A，可从数据集中采样多个与其不同的样本用于构造训练标签，从而扩充训练集，实现数据增强. 本

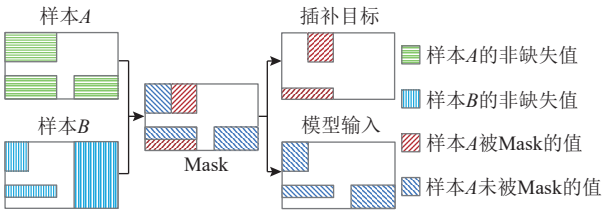


Fig. 3　Illustration of Mask strategy

图 3　Mask 策略示意图

文对每个样本采样了2个与其不同的样本,即将训练集的规模扩大了1倍.

3.4.2 输出模块与损失函数

在得到不确定性指导的时空注意力模块的输出$\boldsymbol{\mu}^{(L)}$和$\boldsymbol{\sigma}^{(L)^2}$后,通过单层前馈网络得到模型输出:

$$\begin{aligned}\boldsymbol{\mu} &= relu(\boldsymbol{\mu}^{(L)}\boldsymbol{W}_{\mathrm{O}}^{\mu}+\boldsymbol{b}_{\mathrm{O}}^{\mu}),\\ \boldsymbol{\sigma}^2 &= softplus(\boldsymbol{\sigma}^{(L)^2}\boldsymbol{W}_{\mathrm{O}}^{\sigma}+\boldsymbol{b}_{\mathrm{O}}^{\sigma}),\end{aligned} \tag{17}$$

其中$\boldsymbol{W}_{\mathrm{O}}^{\mu}$, $\boldsymbol{W}_{\mathrm{O}}^{\sigma}$, $\boldsymbol{b}_{\mathrm{O}}^{\mu}$, $\boldsymbol{b}_{\mathrm{O}}^{\sigma}$为可学习参数.

设B为掩码矩阵的元素和,由于本文假设$\boldsymbol{\mu}$和$\boldsymbol{\sigma}^2$是同一个高斯分布的参数,因此可采用负对数似然损失函数:

$$\mathcal{L}_{\mathrm{nll}}=\frac{1}{B}\sum_{v,t}\left[\frac{(x_{v,t}-\mu_{v,t})^2}{2\sigma_{v,t}^2}+\frac{\ln\sigma_{v,t}^2}{2}\right]m_{v,t}. \tag{18}$$

然而,负对数似然损失函数存在2个问题:1)未给方差指定明确的学习标签;2)方差同时出现在了分母和对数中,难以优化.因此本文额外使用了另一种损失函数:

$$\mathcal{L}_{\mathrm{cons}}=\frac{1}{B}\sum_{v,t}|x_{v,t}-\mu_{v,t}|m_{v,t}+\frac{1}{B}\sum_{v,t}|\gamma|x_{v,t}-\mu_{v,t}|-\sigma_{v,t}|m_{v,t}, \tag{19}$$

其中γ为超参数,该损失函数能解决负对数似然损失函数的问题.最终的损失函数定义为

$$\mathcal{L}=\mathcal{L}_{\mathrm{nll}}+\kappa\mathcal{L}_{\mathrm{cons}}, \tag{20}$$

其中κ为超参数.虽然本文在计算损失函数时仅考虑了非缺失值,但由于采用Mask策略删除了一部分非缺失值,因此可以让模型学习如何复原缺失值.

4 实 验

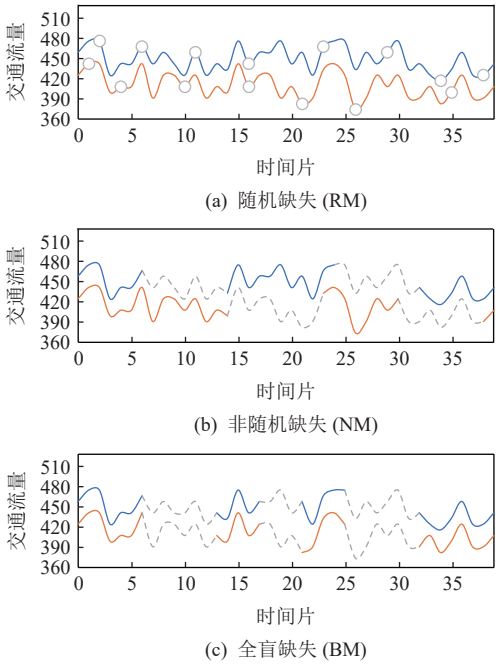
本节将在2个交通流量数据集上评估STUIN的性能,并对STUIN的重要模块和参数进行分析.

4.1 实验设置

4.1.1 数据集

本文使用交通流量数据集PEMS04和PEMS08[4]进行实验,它们的采集频率都是每30秒1次,本文按照每5分钟1个时间片进行聚合.其中PEMS04数据集来自旧金山湾区,时间跨度为2018年1月1日到2018年2月28日,包含307个传感器.PEMS08数据集来自圣贝纳迪诺地区,时间跨度为2016年7月1日到2016年8月31日,包含170个传感器.

本文在这2种数据集上构造缺失值,在测试时针对构造的缺失值计算评价指标,对比不同模型的性能.缺失模式包括3种:随机缺失(random missing, RM)、非随机缺失(non-random missing, NM)和全盲缺失(blackout missing, BM)[9],缺失率为20%, 30%, 70%, 90%. RM指数据完全随机缺失;NM指缺失值在时间维度上是连续出现的;BM指所有节点的数据在一段时间内全部缺失.其中NM模式下的缺失值在时间维度上具有相关性,BM模式下的缺失值在时间和空间维度上都有相关性.这3种缺失模式的示意图如图4所示,其中虚线部分表示缺失.



(a) 随机缺失 (RM)

(b) 非随机缺失 (NM)

(c) 全盲缺失 (BM)

注:交通流量指每5分钟通过的车辆数.

Fig. 4 Illustration of miss patterns

图4 缺失模式示意图

4.1.2 评价指标

本文使用平均绝对误差(mean absolute error, *MAE*)、均方根误差(root mean square error, *RMSE*)和平均绝对百分比误差(mean absolute percentage error, *MAPE*)衡量插补性能,设y_i代表真实值,$\hat{y}_i$代表预测值,具体计算公式为:

$$\begin{aligned}MAE &= \frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i|,\\ RMSE &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2},\\ MAPE &= \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i-\hat{y}_i|}{|y_i|}.\end{aligned} \tag{21}$$

使用预测区间覆盖率(prediction interval coverage probability, *PICP*)和平均间隔分数(mean interval score, *MIS*)衡量不确定性量化性能.设置信度为$1-\rho$, $\rho/2$

在标准正态分布表中对应的z值为λ，则可以根据标准差计算上下界，具体公式为

$$\begin{aligned} l_i &= \hat{y}_i - \lambda\sigma_i, \\ u_i &= \hat{y}_i + \lambda\sigma_i. \end{aligned} \tag{22}$$

使用 *PICP* 衡量置信区间包含真实值的概率：

$$PICP = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{I}(l_i < y_i < u_i), \tag{23}$$

PICP 越大说明真实值落入置信区间的概率越大. 然而，若置信区间范围过大，*PICP* 也会较高，但这没有意义，因此使用 *MIS* 衡量置信区间是否足够紧致：

$$\begin{aligned} MIS = &\frac{1}{N}\sum_{i=1}^{N}\left\{(u_i - l_i) + \frac{2}{\rho}(y_i - u_i)\boldsymbol{I}(y_i > u_i) + \right. \\ &\left. \frac{2}{\rho}(l_i - y_i)\boldsymbol{I}(l_i > y_i)\right\}, \end{aligned} \tag{24}$$

在本文中$\rho = 0.05$，其对应的$\lambda = 1.96$.

4.1.3 基准模型

本文使用 5 种具有代表性的插补模型和 STUIN 对比，评估插补性能.

1）LATC[9]. 适用于时空交通数据的低秩自回归张量补全框架.

2）MIWAE[10]. 基于重要性加权自编码器的数据插补模型.

3）BRITS[12]. 基于双向循环神经网络的时间序列插补模型.

4）mTAN[13]. 处理不规则采样时间序列的表示学习模型，能用于实现数据插补.

5）IGNNK[14]. 适用于时空交通数据的克里金插值模型.

将 STUIN 与下面 5 种不确定性量化模型对比，评估不确定性量化性能.

1）MC Dropout[24]. 通过随机 Dropout 实现变分推断，在测试时开启 Dropout，使用多次预测的统计数据量化不确定性.

2）平均间隔分数回归（MIS-Reg）[26]. 通过最小化 *MIS* 估计置信区间.

3）分位数回归（Quan-Reg）[27]. 通过最小化单边分位数损失函数估计置信区间.

4）样条分位数回归（SQ-Reg）[28]. 通过最小化 CRPS 拟合分位数函数，实现不确定性量化.

5）DeepAR[29]. 基于自回归循环神经网络的概率预测模型.

其中 MC Dropout、MIS-Reg、Quan-Reg 和 SQ-Reg 都基于 DCRNN[15] 实现.

4.1.4 参数设置

本文通过 Min-Max 归一化将数据标准化到[−1, 1]，将缺失数据填充为 0. 之后将数据集按照 6∶2∶2 的比例划分为训练集、验证集和测试集，并在 3 个数据集上分别使用长度为 13 的滑动窗口构造样本，然后在训练集和验证集上使用 Mask 策略构造训练标签. 参数设置为$L = 6$, $d = 64$, $\tau = 1.0$, $\nu = 1.0$, $\gamma = 2$, $\kappa = 1.5$，注意力头数为 8，使用 PyTorch 框架实现模型，学习率为 0.001，批量大小为 16.

4.2 性能比较

由于缺失模式有 3 种，缺失率有 4 种，共有 12 种实验设置，数据量过大，因此正文中仅展示了 PEMS04 数据集上 RM 缺失模式的实验结果，如表 1 和表 2 所示，其他实验结果见附录 A. 可以看出，在 RM 缺失模式下，STUIN 在插补指标上均优于基准模型，并且随着缺失率的增加优势越来越明显，因此其插补性能非常出色. 在不确定性量化指标上，STUIN 预测的置信区间不但具有最高的区间覆盖率，并且 *MIS* 也是所有模型中最小的. 因此，STUIN 能够提供准确而紧致的置信区间. 在其他数据集上的表现与此类似.

在插补模型中，BRITS 和 mTAN 适用于时间序列数据，当缺失率较低时在时空图数据上也具有良好的性能，反之性能明显下降. MIWAE 适用于服从独立同分布假设的数据集，虽然大部分情况下性能较差，但在 90% 缺失率下是基准模型中最优的. 这可能是因为其模型结构较为简单，当缺失率过高时更不容易发生过拟合. LATC 是传统机器学习模型，在缺失率较低时具有良好的性能，但缺失率过高时性能明显下降. IGNNK 虽然适用于时空图数据，但针对的是克里金插值，且对时间相关性的处理过于简单，整体性能一般. 在不确定性量化模型中，MC Dropout 在大部分情况下都是基准模型中最优的，这是因为 MC Dropout 在预测时相当于一个集成模型. MIS-Reg、Quan-Reg 和 SQ-Reg 的性能都随着缺失率的升高迅速下降. DeepAR 在低缺失率下具有良好的插补性能，但当缺失率为 90% 时不再具有优势.

4.3 可视化分析

在本节中，我们对 STUIN 的不确定性量化性能进行分析. 使用的数据集为 PEMS04，缺失模式为 NM，缺失率为 70%. 在测试集中随机选择某节点的一小段数据，对 STUIN、MC Dropout、MIS-Reg 和 DeepAR 输出的置信区间进行了可视化，由图 5 可知，STUIN 输出的置信区间几乎在所有情况下都能覆盖真实值，而 MC Dropout、MIS-Reg 和 DeepAR 都存在真实值不

Table 1 Performance Comparison of 20% and 30% Missing Rates in RM Missing Pattern on PEMS04 Dataset

表 1 PEMS04 数据集上 RM 缺失模式下 20% 和 30% 缺失率的性能对比

模型	20%, RM					30%, RM				
	MAE	*RMSE*	*MAPE*/%	*PICP*/%	*MIS*	*MAE*	*RMSE*	*MAPE*/%	*PICP*/%	*MIS*
LATC[9]	19.63	31.70	15.16			20.08	32.48	15.61		
MIWAE[10]	29.22	48.88	19.78			29.56	49.40	19.69		
BRITS[12]	21.13	35.50	26.88			20.51	34.31	24.03		
mTAN[13]	18.36	29.66	13.32			19.27	31.31	13.25		
IGNNK[14]	22.11	34.33	19.41			21.25	33.29	21.22		
MC Dropout[24]	15.89	26.31	12.24	70.54	220.95	16.32	26.93	12.76	70.89	224.75
MIS-Reg[26]	18.56	29.91	14.02	62.06	285.11	21.65	34.90	16.17	53.75	414.17
Quan-Reg[27]	21.96	35.54	17.66	67.87	230.17	23.00	35.97	19.39	45.72	443.38
SQ-Reg[28]	18.12	29.33	13.71	64.48	255.49	18.36	29.87	13.81	62.79	277.04
DeepAR[29]	19.67	31.67	12.60	74.74	130.44	20.17	32.26	13.32	68.73	148.37
STUIN（本文）	**15.55**	**25.67**	**10.81**	**97.71**	**109.57**	**15.55**	**25.39**	**11.07**	**97.19**	**107.48**

注：下划线数值表示次优指标，黑体数值表示最优指标.

Table 2 Performance Comparison of 70% and 90% Missing Rates in RM Missing Pattern on PEMS04 Dataset

表 2 PEMS04 数据集上 RM 缺失模式下 70% 和 90% 缺失率的性能对比

模型	70%, RM					90%, RM				
	MAE	*RMSE*	*MAPE*/%	*PICP*/%	*MIS*	*MAE*	*RMSE*	*MAPE*/%	*PICP*/%	*MIS*
LATC[9]	28.76	45.38	21.37			61.60	87.36	67.58		
MIWAE[10]	31.67	51.54	20.98			34.70	55.10	23.41		
BRITS[12]	26.09	40.19	22.93			45.20	67.12	37.52		
mTAN[13]	24.21	41.99	18.18			58.53	100.48	66.81		
IGNNK[14]	24.18	38.30	22.37			42.05	65.55	43.60		
MC Dropout[24]	21.81	41.82	16.24	70.74	344.44	63.42	118.19	37.16	51.20	1 762.39
MIS-Reg[26]	28.51	49.68	20.25	37.00	780.89	73.89	125.72	44.18	56.63	1 888.09
Quan-Reg[27]	31.23	52.14	23.17	31.41	959.31	73.13	130.63	43.06	18.85	2 741.20
SQ-Reg[28]	29.29	51.36	21.03	43.16	727.01	79.26	133.08	54.28	52.11	1 984.74
DeepAR[29]	29.14	53.18	16.83	50.22	316.94	104.96	163.99	52.66	17.63	1 932.63
STUIN（本文）	**18.05**	**30.77**	**12.52**	**96.97**	**129.23**	**23.67**	**37.33**	**18.85**	**93.01**	**177.18**

注：下划线数值表示次优指标，黑体数值表示最优指标.

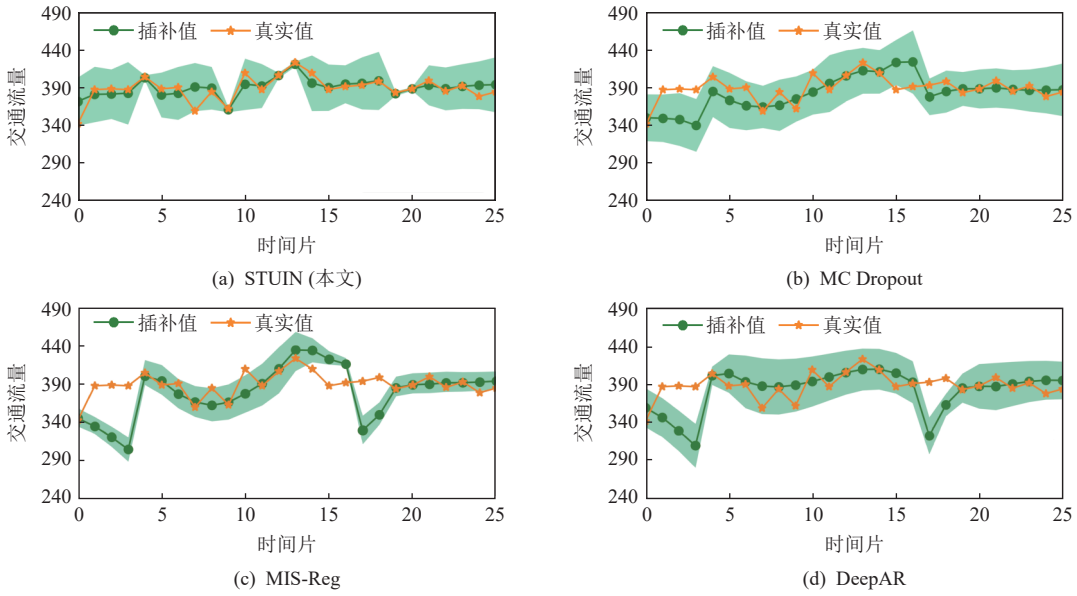
在置信区间内的情况. 此外, STUIN 的置信区间也较为紧致. 因此, STUIN 具有出色的不确定性量化性能.

4.4 超参数分析

本节将对损失函数中的 2 个重要超参数: γ和κ进行分析. 使用的数据集为 PEMS04, 缺失模式为 RM, 缺失率为 20%. 首先分析超参数γ对性能的影响, 固定$\kappa=1.5$, 分别将γ设置为 1, 2, 3 进行实验, 实验结果如图 6 所示. 由图 6 可知, 随着γ的增大 *MAE* 和 *MAPE* 逐渐减小, 这说明增大γ能提高插补性能. 当$\gamma=1$时 *PICP* 和 *MIS* 指标均较差; 当$\gamma=2$时 *PICP* 和 *MIS* 达到图中的最优; 当$\gamma=3$时虽然 *PICP* 有所提升, 但 *MIS* 也显著升高, 因为此时方差必须足够大才能使损失函数降低. 因此, 尽管增大γ能提升插补性能, 但γ过大会降低置信区间的预测质量.

其次分析超参数κ的影响, 固定$\gamma=2$, 分别将κ设置为 0, 1, 2 进行实验, 实验结果如图 7 所示. 由图 7 可知, 随着κ的增大, 插补性能得到提升, 尤其是κ从 0 增加至 1 时性能提升显著, 这是因为$\kappa=0$时损失函数仅包含负对数似然. 而κ从 1 增加至 2 时性能并没有太大提升. *PICP* 和 *MIS* 具有类似的表现.

最后, 本节对参数的鲁棒性进行分析. 将上述不同参数设置下得到的实验结果与 MC Dropout 进行对



注：交通流量指每 5 分钟通过的车辆数.

Fig. 5 Confidence interval visualization

图 5 置信区间可视化

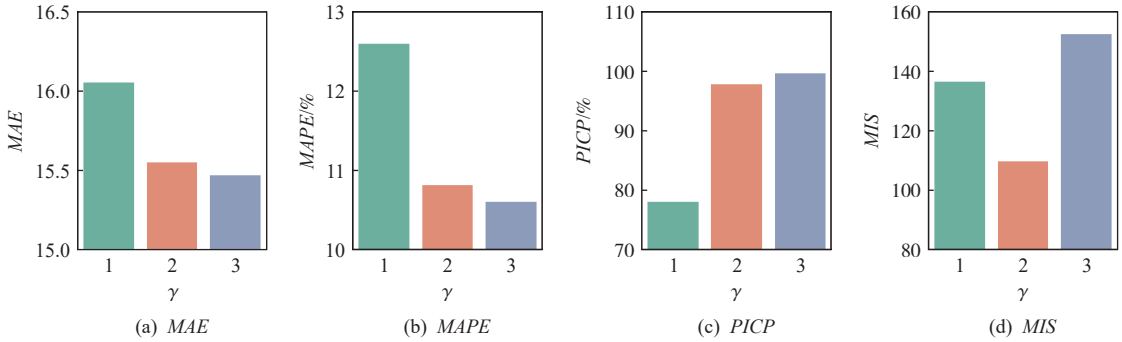


Fig. 6 Analysis of hyper-parameter γ

图 6 γ超参数分析

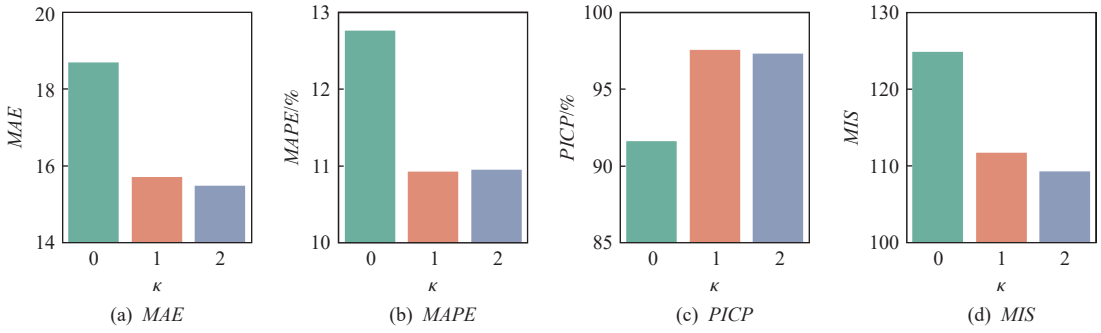


Fig. 7 Analysis of hyper-parameter κ

图 7 κ超参数分析

比，MC Dropout 的整体性能是基准模型中最优的，实验结果如表 3 所示. 可以看出，在绝大部分参数设置下，STUIN 的性能都显著优于 MC Dropout，即便是在$\gamma=1$或$\kappa=0$这种较极端的取值情况下，STUIN 的不确定性量化性能也显著领先于 MC Dropout，因此 STUIN 对参数取值不敏感，具有较强的鲁棒性.

4.5 消融实验

本节将通过消融实验证明 STUIN 各组件的有效

Table 3 Robustness Analysis

表 3 鲁棒性分析

模型		*MAE*	*MAPE*/%	*PICP*/%	*MIS*
MC Dropout[24]		15.89	12.24	70.54	220.95
STUIN（本文）	$\kappa=1.5,\gamma=1$	16.05	12.59	78.06	136.39
	$\kappa=1.5,\gamma=2$	15.55	10.81	97.71	109.57
	$\kappa=1.5,\gamma=3$	15.47	10.60	99.58	152.36
	$\gamma=2,\kappa=0$	18.68	12.76	91.61	124.82
	$\gamma=2,\kappa=1$	15.71	10.93	97.55	111.70
	$\gamma=2,\kappa=2$	15.48	10.95	97.31	109.31

性. 分别移除 STUIN 的重要组件, 并将实验结果与完整的 STUIN 进行对比. STUIN 包括 3 个变种:

1)noSE. 移除时空不确定性初始化模块中的空间嵌入矩阵.

2)noTE. 移除时空不确定性初始化模块中的时间嵌入矩阵.

3)noVA. 不再将隐状态定义成高斯分布, 即移除方差, 此时模型仅能给出点估计.

数据集为 PEMS04, 缺失模式为 RM, 缺失率为 20%, 变种模型的参数设置和 STUIN 一致, 实验结果如表 4 所示.

Table 4 Ablation Experiments

表 4 消融实验

模型	*MAE*	*MAPE*/%	*PICP*/%	*MIS*
noSE	15.92	11.55	95.73	111.81
noTE	15.71	11.69	96.48	110.86
noVA	15.97	11.32		
STUIN（本文）	15.55	10.81	97.71	109.57

从表 4 可以看出, 当移除不确定性建模时, 数据插补性能明显下降. 这说明本文提出的不确定性指导的时空依赖性学习机制能起到削弱无效信息传播的作用. 同时, 移除时间嵌入矩阵和空间嵌入矩阵会导致插补和不确定性量化性能都下降. 这说明同时考虑时间和空间因素能更准确地量化不确定性, 进而提高不确定性在学习时空依赖性时的指导作用.

4.6 下游任务实验

本节将通过实验验证对缺失数据进行补全能提升下游任务的精度. 使用的数据集为 PEMS04, 缺失率为 20%, 涵盖所有缺失模式. 在训练完模型后, 使用模型输出的插补值补全缺失值, 保留非缺失值:

$$\boldsymbol{X}=\boldsymbol{X}\odot\boldsymbol{M}+(1-\boldsymbol{M})\odot\boldsymbol{\mu}. \tag{25}$$

然后使用这些数据集训练时空预测模型 AGCRN[33], 并在相同的测试集上计算评价指标(测试集来自原始 PEMS04 数据集). AGCRN 的参数设置和原文一致, 实验结果如表 5 所示.

Table 5 Downstream Task Experiments on PEMS04 Dataset

表 5 PEMS04 数据集上的下游任务实验

数据集		*MAE*	*RMSE*	*MAPE*/%
无缺失		19.84	32.52	12.94
RM	缺失	21.45	34.97	14.81
	补全	19.62	32.12	13.02
NM	缺失	21.71	34.82	15.44
	补全	19.76	32.23	12.81
BM	缺失	20.06	33.13	13.12
	补全	19.74	32.42	13.12

由表 5 可以看出, 无论是哪种缺失模式, 经过数据补全后都可以明显提升时空预测的性能, 这说明对缺失交通数据进行补全能显著提高下游任务的精度, 具有重要的应用价值.

5 总结与展望

本文提出了一种时空不确定性指导的交通数据插补模型 STUIN, 能同时为缺失值提供点估计和置信区间. STUIN 将神经网络的隐状态建模为服从高斯分布的随机变量, 使用方差建模隐状态的不确定性, 利用基于方差的注意力机制指导模型学习时空相关性. 此外, 本文还提出了一个新颖的时空不确定性初始化模块, 在初始化均值和方差时同时考虑了时间、空间和数据缺失状况. 最后, 本文在 2 个交通流量数据集上构造了多种缺失模式, 设置了不同的缺失率进行实验, 实验结果表明, STUIN 具有出色的数据插补和不确定性量化性能.

在未来, 我们将从 2 个方面开展进一步研究: 首先, 本文提出的模型通过 Mask 非缺失值构造训练标签, 实现自监督学习, 而自监督学习领域目前已发展出了更为有效的策略, 如对比学习等, 未来我们将尝试将这些技术应用于模型. 其次, 由于方差嵌入函数的存在, 本文提出的时空不确定性初始化模块只能应用于单一特征, 未来我们将针对多维特征数据集的插补和不确定性量化开展研究.

作者贡献声明: 刘乐提出了算法思路和实验方案, 负责完成实验并撰写论文; 郭晟楠、靳希源、林友

芳和万怀宇对模型的设计提出指导意见并进行了论文的修改；赵苗苗和陈冉辅助做了实验结果的统计和分析.

参 考 文 献

[1] Qu Li, Li Li, Zhang Yi, et al. PPCA-based missing data imputation for traffic flow volume: A systematical approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2009, 10(3): 512−522

[2] Abdar M, Pourpanah F, Hussain S, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges[J]. Information Fusion, 2021, 76: 243-297

[3] Guo Shengnan, Lin Youfang, Wan Huaiyu, et al. Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(11): 5415−5428

[4] Guo Shengnan, Lin Youfang, Feng Ning, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[C]//Proc of the 33rd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2019: 922−929

[5] Yuan Haitao, Li Guoliang. A survey of traffic prediction: From spatio-temporal data to intelligent transportation[J]. Data Science and Engineering, 2021, 6: 63−85

[6] Ni Qingjian, Peng Wenqiang, Zhang Zhizheng, et al. Spatial-temporal graph neural network for traffic flow prediction based on information enhanced transmission[J]. Journal of Computer Research and Development, 2022, 59(2): 282−293 (in Chinese)
(倪庆剑, 彭文强, 张志政, 等. 基于信息增强传输的时空图神经网络交通流预测[J]. 计算机研究与发展, 2022, 59(2): 282−293)

[7] Lebel J M, Kratz F, Bloch G. Missing values rebuilding by prediction and ARIMA modelling in time series[J]. Advanced Information Processing in Automatic Control, 1989, 22(6): 357−361

[8] Hudak A T, Crookston N L, Evans J S, et al. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from lidar data[J]. Remote Sensing of Environment, 2008, 112(5): 2232−2245

[9] Chen Xinyu, Lei Mengying, Saunier N, et al. Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(8): 12301−12310

[10] Mattei P A, Frellsen J. MIWAE: Deep generative modelling and imputation of incomplete data sets[C]//Proc of the 36th Int Conf on Machine Learning. New York: ACM, 2019: 4413−4423

[11] Che Zhengping, Purushotham S, Cho K, et al. Recurrent neural networks for multivariate time series with missing values[J]. Scientific Reports, 2018, 8(1): 6085−6099

[12] Cao Wei, Wang Dong, Li Jian, et al. BRITS: Bidirectional recurrent imputation for time series[C]//Proc of the 32nd Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2018: 6776−6786

[13] Shukla S N, Marlin B M. Multi-time attention networks for irregularly sampled time series[C/OL]//Proc of the 9th Int Conf on Learning Representations. San Juan, CA: ICLR, 2021[2021-11-21]. https://openreview.net/forum?id=mXbhcalKnYM

[14] Wu Yuankai, Zhuang Dingyi, Labbe A, et al. Inductive graph neural networks for spatiotemporal Kriging[C]//Proc of the 35th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2021: 4478−4485

[15] Li Yaguang, Yu R, Shahabi C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting[C/OL]//Proc of the 6th Int Conf on Learning Representations. San Juan, CA: ICLR, 2018[2021-08-04]. https://openreview.net/forum?id=SJiHXGWAZ

[16] Xu Dongwei, Peng Hang, Wei Chenchen, et al. Traffic state data imputation: An efficient generating method based on the graph aggregator[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(8): 13084−13093

[17] Hamilton W, Ying Zhitao, Leskovec J. Inductive representation learning on large graphs[C]//Proc of the 31st Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2017: 1024−1034

[18] Ye Yongchao, Zhang Shiyao, Yu J J Q. Spatial-temporal traffic data imputation via graph attention convolutional network[C]//Proc of the 30th Int Conf on Artificial Neural Networks. Berlin: Springer, 2021: 241−252

[19] Blundell C, Cornebise J, Kavukcuoglu K, et al. Weight uncertainty in neural networks[C]//Proc of the 32nd Int Conf on Machine Learning. New York: ACM, 2015: 1613−1622

[20] Louizos C, Welling M. Multiplicative normalizing flows for variational Bayesian neural networks[C]//Proc of the 34th Int Conf on Machine Learning. New York: ACM, 2017: 2218–2227

[21] Wu Anqi, Nowozin S, Meeds E, et al. Deterministic variational inference for robust Bayesian neural networks[C/OL]//Proc of the 7th Int Conf on Learning Representations. San Juan, CA: ICLR, 2019[2022-03-14]. https://openreview.net/forum?id=B1l08oAct7

[22] Welling M, Teh Y W. Bayesian learning via stochastic gradient Langevin dynamics[C]//Proc of the 28th Int Conf on Machine Learning. New York: ACM, 2011: 681−688

[23] Li Chunyuan, Chen Changyou, Carlson D E, et al. Preconditioned stochastic gradient Langevin dynamics for deep neural networks[C]//Proc of the 30th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2016: 1788−1794

[24] Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning[C]//Proc of the 33rd Int Conf on Machine Learning. New York: ACM, 2016: 1050−1059

[25] Gal Y, Ghahramani Z. A theoretically grounded application of dropout in recurrent neural networks[C]//Proc of the 30th Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2016: 1027−1035

[26] Wu Dongxia, Gao Liyao, Chinazzi M, et al. Quantifying uncertainty in deep spatial-temporal forecasting[C]//Proc of the 27th ACM SIGKDD Conf on Knowledge Discovery and Data Mining. New York: ACM, 2021: 1841−1851

[27] Koenker R, Hallock K. Quantile regression[J]. Journal of Economic Perspectives, 2001, 15(4): 143−156

[28] Gasthaus J, Benidis K, Wang Yuyang, et al. Probabilistic forecasting with spline quantile function RNNs[C]//Proc of the 22nd Int Conf on Artificial Intelligence and Statistics. Cambridge, MA: JMLR, 2019: 1901−1910

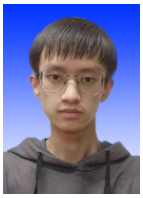
[29] Salinas D, Flunkert V, Gasthaus J, et al. DeepAR: Probabilistic forecasting with autoregressive recurrent networks[J]. International Journal of Forecasting, 2020, 36(3): 1181−1191

[30] Zhou Zhengyang, Wang Yang, Xie Xike, et al. STUaNet: Understanding uncertainty in spatiotemporal collective human mobility[C]//Proc of the 30th Int World Wide Web Conf. Berlin: Springer, 2021: 1868−1879

[31] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proc of the 31st Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2017: 6000−6010

[32] Tashiro Y, Song Jiaming, Song Yang, et al. CSDI: Conditional score-based diffusion models for probabilistic time series imputation[C]//Proc of the 35th Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2021: 24804−24816

[33] Bai Lei, Yao Lina, Li Can, et al. Adaptive graph convolutional recurrent network for traffic forecasting[C]//Proc of the 34th Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2020: 17804−17815



Liu Le, born in 1999. Master. His main research interests include spatial-temporal data mining and deep learning.

刘　乐, 1999 年生. 硕士. 主要研究方向为时空数据挖掘、深度学习.



Guo Shengnan, born in 1992. PhD, lecturer. Member of CCF. Her main research interests include spatial-temporal data mining and deep learning.

郭晟楠, 1992 年生. 博士, 讲师. CCF 会员. 主要研究方向为时空数据挖掘、深度学习.



Jin Xiyuan, born in 1997. PhD candidate. Student member of CCF. His main research interests include spatial-temporal data mining and uncertainty estimation.

靳希源, 1997 年生. 博士研究生. CCF 学生会员. 主要研究方向为时空数据挖掘、不确定性量化.



Zhao Miaomiao, born in 2002. Undergraduate. Her main research interest includes spatial-temporal data mining.

赵苗苗, 2002 年生. 本科生. 主要研究方向为时空数据挖掘.



Chen Ran, born in 2002. Undergraduate. Her main research interest includes spatial-temporal data imputation.

陈　冉, 2002 年生. 本科生. 主要研究方向为时空数据插补.



Lin Youfang, born in 1971. PhD, professor. Senior member of CCF. His main research interests include data mining, machine learning, reinforcement learning, complex network, and intelligent technology and system.

林友芳, 1971 年生. 博士, 教授. CCF 高级会员. 主要研究方向为数据挖掘、机器学习、强化学习、复杂网络、智能技术与系统.



Wan Huaiyu, born in 1981. PhD, professor. Member of CCF. His main research interests include spatial-temporal data mining, information extraction、social networks mining.

万怀宇, 1981 年生. 博士, 教授. CCF 会员. 主要研究方向为时空数据挖掘、信息抽取、社交网络挖掘.

附录 A.

Table A1 Performance Comparison of 20% and 30% Missing Rates in NM Missing Pattern on PEMS04 Dataset

表 A1 PEMS04 数据集上 NM 缺失模式下 20% 和 30% 缺失率的性能对比

模型	20%, NM					30%, NM				
	MAE	*RMSE*	*MAPE/%/*	*PICP/%*	*MIS*	*MAE*	*RMSE*	*MAPE/%*	*PICP/%*	*MIS*
LATC[9]	40.34	66.31	21.87			62.58	95.91	32.00		
MIWAE[10]	29.42	49.01	19.29			28.78	48.36	19.04		
BRITS[12]	23.19	37.71	28.24			24.11	37.15	31.22		
mTAN[13]	19.19	30.95	13.92			19.17	31.14	13.38		
IGNNK[14]	22.19	34.47	21.08			21.54	33.25	21.54		
MC Dropout[24]	16.20	26.67	12.58	71.90	219.16	16.45	26.94	12.58	70.89	226.92
MIS-Reg[26]	20.14	32.37	14.70	58.33	344.30	19.75	31.47	15.33	55.77	343.44
Quan-Reg[27]	24.87	38.13	20.40	57.82	355.23	23.54	36.17	19.95	51.74	399.72
SQ-Reg[28]	18.33	29.67	13.70	67.82	244.05	19.24	31.07	14.77	60.01	312.51
DeepAR[29]	20.34	32.30	13.13	62.73	159.74	21.90	34.17	14.78	60.12	181.46
STUIN（本文）	**15.68**	**25.82**	**10.91**	**96.69**	**110.49**	**15.64**	**25.80**	**10.91**	**96.16**	**109.53**

注：下划线数值表示次优指标，黑体数值表示最优指标.

Table A2 Performance Comparison of 70% and 90% Missing Rates in NM Missing Pattern on PEMS04 Dataset

表 A2 PEMS04 数据集上 NM 缺失模式下 70% 和 90% 缺失率的性能对比

模型	70%, NM					90%, NM				
	MAE	*RMSE*	*MAPE/%*	*PICP/%*	*MIS*	*MAE*	*RMSE*	*MAPE/%*	*PICP/%*	*MIS*
LATC[9]	138.64	187.81	67.67			176.56	228.98	82.82		
MIWAE[10]	29.53	49.08	19.32			30.17	49.31	19.80		
BRITS[12]	26.50	41.57	28.78			47.66	70.15	36.27		
mTAN[13]	29.60	67.85	21.41			94.91	165.32	61.46		
IGNNK[14]	24.04	37.84	22.02			40.53	60.75	43.73		
MC Dropout[24]	22.21	43.71	16.36	70.51	358.17	57.21	103.62	34.88	53.31	1 355.99
MIS-Reg[26]	30.84	54.34	21.82	40.82	811.53	73.68	129.06	44.18	21.12	2 598.60
Quan-Reg[27]	28.16	49.76	22.15	42.43	686.49	70.66	125.10	42.18	21.91	2 603.93
SQ-Reg[28]	26.28	48.26	19.49	54.79	542.92	74.72	130.42	45.46	25.94	2 537.45
DeepAR[29]	32.82	61.07	17.62	59.67	308.22	103.80	158.26	52.67	28.43	1 654.39
STUIN（本文）	**18.03**	**30.40**	**12.30**	**97.24**	**127.23**	**23.64**	**37.74**	**17.26**	**94.74**	**172.45**

注：下划线数值表示次优指标，黑体数值表示最优指标.

Table A3 Performance Comparison of 20% and 30% Missing Rates in BM Missing Pattern on PEMS04 Dataset

表 A3 PEMS04 数据集上 BM 缺失模式下 20% 和 30% 缺失率的性能对比

模型	20%, BM					30%, BM				
	MAE	*RMSE*	*MAPE/%*	*PICP/%*	*MIS*	*MAE*	*RMSE*	*MAPE/%*	*PICP/%*	*MIS*
LATC[9]	27.07	42.74	20.11			32.84	50.11	23.51		
MIWAE[10]	107.33	133.59	151.68			106.87	133.17	161.59		
BRITS[12]	22.84	37.06	30.31			23.63	37.21	31.90		
mTAN[13]	18.76	30.37	12.97			19.25	31.37	13.25		
IGNNK[14]	33.90	58.34	42.11			40.97	68.58	60.40		
MC Dropout[24]	16.40	26.64	12.97	71.33	222.55	19.23	29.89	13.08	68.25	277.28
MIS-Reg[26]	17.71	28.46	13.65	71.93	207.99	18.75	30.08	13.16	71.51	229.37
Quan-Reg[27]	26.45	39.64	23.42	70.98	222.66	27.69	41.40	22.03	67.16	266.15
SQ-Reg[28]	17.68	28.55	13.87	71.52	206.70	19.06	30.33	13.66	69.09	241.72
DeepAR[29]	20.59	32.90	12.99	58.05	169.12	21.03	33.92	13.76	61.45	187.48
STUIN（本文）	**16.03**	**25.88**	**10.88**	**97.33**	**117.73**	**15.88**	**25.76**	**10.69**	**97.16**	**116.86**

注：下划线数值表示次优指标，黑体数值表示最优指标.

Table A4 Performance Comparison of 70% and 90% Missing Rates in BM Missing Pattern on PEMS04 Dataset

表 A4 PEMS04 数据集上 BM 缺失模式下 70% 和 90% 缺失率的性能对比

模型	70%, BM					90%, BM				
	MAE	*RMSE*	*MAPE/%*	*PICP/%*	*MIS*	*MAE*	*RMSE*	*MAPE/%*	*PICP/%*	*MIS*
LATC[9]	43.46	69.32	35.23			93.16	138.74	55.16		
MIWAE[10]	107.76	134.09	151.74			109.76	136.69	141.26		
BRITS[12]	26.45	40.98	28.53			55.34	78.80	51.44		
mTAN[13]	22.88	39.40	15.76			59.51	102.58	64.93		
IGNNK[14]	82.69	117.31	135.19			121.34	149.73	227.13		
MC Dropout[24]	19.09	31.29	15.10	71.98	261.71	49.11	90.57	46.17	59.43	1 219.17
MIS-Reg[26]	21.31	32.46	19.76	68.55	278.66	49.33	88.17	51.27	24.30	1 649.11
Quan-Reg[27]	27.90	41.89	24.93	58.25	359.07	60.54	108.78	47.81	25.10	2 000.48
SQ-Reg[28]	20.89	33.22	16.71	53.22	393.67	48.41	84.21	57.48	27.61	1 699.75
DeepAR[29]	34.36	53.29	21.45	41.33	368.13	105.36	161.35	53.72	21.32	1 664.47
STUIN（本文）	**18.11**	**30.07**	**12.68**	**97.12**	**130.16**	**43.72**	**73.41**	**46.42**	**97.41**	**331.82**

注：下划线数值表示次优指标，黑体数值表示最优指标.

Table A5 Performance Comparison of 20% and 30% Missing Rates in RM Missing Pattern on PEMS08 Dataset

表 A5 PEMS08 数据集上 RM 缺失模式下 20% 和 30% 缺失率的性能对比

模型	20%, RM					30%, RM				
	MAE	*RMSE*	*MAPE/%*	*PICP/%*	*MIS*	*MAE*	*RMSE*	*MAPE/%*	*PICP/%*	*MIS*
LATC[9]	15.29	24.93	10.93			15.79	25.69	11.32		
MIWAE[10]	27.58	48.03	17.17			28.29	48.85	17.29		
BRITS[12]	14.37	23.49	28.62			15.32	24.46	26.09		
mTAN[13]	13.99	22.02	9.39			18.64	28.99	11.58		
IGNNK[14]	19.43	28.22	31.95			18.49	26.95	25.43		
MC Dropout[24]	13.24	20.45	10.90	74.14	165.76	13.32	20.32	10.61	74.87	162.40
MIS-Reg[26]	14.35	21.73	10.36	74.93	157.75	14.34	21.69	10.46	71.44	177.36
Quan-Reg[27]	21.36	30.17	17.93	72.98	163.17	20.63	29.44	17.10	66.19	209.74
SQ-Reg[28]	14.18	21.55	10.34	76.81	146.13	13.88	21.03	10.20	78.73	131.43
DeepAR[29]	15.66	24.00	9.55	63.66	128.12	17.47	26.31	10.38	48.26	178.01
STUIN（本文）	**10.76**	**18.02**	**7.17**	**97.65**	**79.68**	**11.02**	**18.31**	**7.45**	**97.01**	**78.13**

注：下划线数值表示次优指标，黑体数值表示最优指标.

Table A6 Performance Comparison of 70% and 90% Missing Rates in RM Missing Pattern on PEMS08 Dataset

表 A6 PEMS08 数据集上 RM 缺失模式下 70% 和 90% 缺失率的性能对比

模型	70%, RM					90%, RM				
	MAE	*RMSE*	*MAPE/%*	*PICP/%*	*MIS*	*MAE*	*RMSE*	*MAPE/%*	*PICP/%*	*MIS*
LATC[9]	23.06	36.35	15.65			55.64	77.27	42.45		
MIWAE[10]	30.96	52.33	19.00			33.42	54.31	20.84		
BRITS[12]	21.94	33.91	27.11			49.73	66.54	78.97		
mTAN[13]	18.33	30.24	11.61			57.02	102.32	56.47		
IGNNK[14]	22.52	33.67	31.02			42.69	64.91	69.45		
MC Dropout[24]	16.36	26.76	13.10	75.83	200.11	25.79	42.11	19.40	69.52	378.16
MIS-Reg[26]	19.87	36.77	13.33	51.55	453.80	71.66	128.91	36.08	25.11	2 512.05
Quan-Reg[27]	22.92	38.58	19.08	46.16	506.56	65.51	118.03	35.76	25.92	2 402.22
SQ-Reg[28]	19.02	34.48	13.48	61.24	356.77	65.95	117.28	37.06	52.41	2 040.53
DeepAR[29]	35.76	61.80	18.72	33.84	389.42	103.12	158.41	47.59	14.34	1 726.22
STUIN（本文）	**13.36**	**21.97**	**9.23**	**97.71**	**94.68**	**19.58**	**30.72**	**12.90**	**93.51**	**152.65**

注：下划线数值表示次优指标，黑体数值表示最优指标.

Table A7 Performance Comparison of 20% and 30% Missing Rates in NM Missing Pattern on PEMS08 Dataset

表 A7 PEMS08 数据集上 NM 缺失模式下 20% 和 30% 缺失率的性能对比

模型	20%, NM					30%, NM				
	MAE	*RMSE*	*MAPE/%*	*PICP/%*	*MIS*	*MAE*	*RMSE*	*MAPE/%*	*PICP/%*	*MIS*
LATC[9]	27.62	55.07	15.22			45.34	77.24	20.85		
MIWAE[10]	26.69	46.97	16.84			26.79	46.94	16.67		
BRITS[12]	16.97	26.39	35.88			16.99	26.44	34.97		
mTAN[13]	14.51	22.45	9.82			14.91	23.35	10.01		
IGNNK[14]	18.87	27.22	27.48			19.15	27.69	31.02		
MC Dropout[24]	13.40	20.42	10.86	74.85	164.61	13.32	20.36	11.28	75.51	162.77
MIS-Reg[26]	14.20	21.49	10.54	76.72	152.77	14.28	21.63	10.71	74.18	160.48
Quan-Reg[27]	22.00	30.68	18.13	75.92	152.74	21.76	30.54	19.26	71.32	182.64
SQ-Reg[28]	13.98	21.03	10.38	77.79	141.12	14.59	22.10	11.10	76.75	155.58
DeepAR[29]	16.28	24.86	10.50	58.24	145.35	16.30	24.85	10.24	54.29	157.32
STUIN（本文）	**10.96**	**18.00**	**7.48**	**97.39**	**76.32**	**11.17**	**18.31**	**8.10**	**94.64**	**78.55**

注：下划线数值表示次优指标，黑体数值表示最优指标.

Table A8 Performance Comparison of 70% and 90% Missing Rates in NM Missing Pattern on PEMS08 Dataset

表 A8 PEMS08 数据集上 NM 缺失模式下 70% 和 90% 缺失率的性能对比

模型	70%, NM					90%, NM				
	MAE	*RMSE*	*MAPE/%*	*PICP/%*	*MIS*	*MAE*	*RMSE*	*MAPE/%*	*PICP/%*	*MIS*
LATC[9]	139.98	185.51	58.68			186.56	232.78	77.93		
MIWAE[10]	28.65	49.89	17.36			29.88	50.61	18.12		
BRITS[12]	20.34	31.78	20.93			54.49	75.52	61.27		
mTAN[13]	22.22	48.57	13.22			119.13	206.23	80.74		
IGNNK[14]	22.90	33.51	31.41			38.90	56.74	57.30		
MC Dropout[24]	16.05	25.21	12.80	75.95	192.79	29.24	49.96	21.83	67.33	494.84
MIS-Reg[26]	20.33	38.66	13.80	57.44	422.88	65.46	117.56	34.21	29.75	2 263.82
Quan-Reg[27]	24.20	40.01	19.86	57.02	433.54	61.82	109.21	35.11	31.65	2 170.33
SQ-Reg[28]	19.53	36.24	13.68	67.19	335.97	59.40	103.57	38.88	57.67	1 693.64
DeepAR[29]	34.73	60.17	18.33	38.97	382.47	104.07	158.45	48.46	24.57	1 614.63
STUIN（本文）	**13.34**	**21.79**	**9.51**	**96.02**	**95.66**	**19.88**	**30.88**	**14.04**	**94.46**	**147.84**

注：下划线数值表示次优指标，黑体数值表示最优指标.

Table A9 Performance Comparison of 20% and 30% Missing Rates in BM Missing Pattern on PEMS08 Dataset

表 A9 PEMS08 数据集上 BM 缺失模式下 20% 和 30% 缺失率的性能对比

模型	20%, BM					30%, BM				
	MAE	*RMSE*	*MAPE*/%	*PICP*/%	*MIS*	*MAE*	*RMSE*	*MAPE*/%	*PICP*/%	*MIS*
LATC[9]	22.24	34.26	14.40			28.87	43.79	18.13		
MIWAE[10]	93.06	116.77	82.30			91.87	116.40	85.99		
BRITS[12]	15.99	25.25	30.43			14.72	23.69	27.18		
mTAN[13]	14.22	22.29	9.00			14.51	22.96	9.25		
IGNNK[14]	23.59	33.14	53.04			26.86	37.35	66.18		
MC Dropout[24]	14.12	21.29	12.87	74.41	176.52	14.70	22.20	12.62	73.87	187.32
MIS-Reg[26]	13.95	21.10	11.51	78.69	132.63	14.47	21.92	11.11	77.98	144.91
Quan-Reg[27]	24.76	33.58	22.42	77.24	153.71	25.50	34.54	22.29	77.12	157.62
SQ-Reg[28]	14.19	21.51	11.42	80.06	132.07	14.70	22.23	11.30	79.02	143.61
DeepAR[29]	16.93	26.01	10.11	62.24	144.53	17.07	26.18	10.15	57.87	155.38
STUIN（本文）	**12.13**	**18.86**	**7.73**	**97.95**	**89.62**	**12.32**	**19.08**	**8.12**	**97.83**	**86.51**

注：下划线数值表示次优指标，黑体数值表示最优指标.

Table A10 Performance Comparison of 70% and 90% Missing Rates in BM Missing Pattern on PEMS08 Dataset

表 A10 PEMS08 数据集上 BM 缺失模式下 70% 和 90% 缺失率的性能对比

模型	70%, BM					90%, BM				
	MAE	*RMSE*	*MAPE*/%	*PICP*/%	*MIS*	*MAE*	*RMSE*	*MAPE*/%	*PICP*/%	*MIS*
LATC[9]	44.26	65.64	32.40			123.35	167.48	58.27		
MIWAE[10]	93.42	117.03	82.93			93.67	117.27	79.88		
BRITS[12]	20.53	31.76	20.65			59.41	80.21	50.67		
mTAN[13]	19.53	32.36	12.69			54.82	91.15	74.97		
IGNNK[14]	27.43	42.58	52.15			58.08	90.21	115.84		
MC Dropout[24]	21.01	40.16	17.20	72.07	337.79	47.65	79.70	61.98	50.71	1 148.35
MIS-Reg[26]	20.88	38.56	17.36	64.09	383.25	45.79	78.75	52.41	59.11	1 069.38
Quan-Reg[27]	33.25	48.18	39.62	74.95	321.66	49.15	84.05	58.61	36.71	1 399.25
SQ-Reg[28]	21.03	37.82	18.37	67.34	358.14	43.87	82.05	46.57	44.30	1 420.40
DeepAR[29]	37.62	67.68	19.92	36.76	453.65	86.40	141.99	40.22	26.98	1 401.37
STUIN（本文）	**16.98**	**29.56**	**12.08**	**97.16**	**121.01**	**35.96**	**60.63**	**30.02**	**96.76**	**288.40**

注：下划线数值表示次优指标，黑体数值表示最优指标.