



计算机科学

COMPUTER SCIENCE

基于缺失数据的交通速度预测算法

黄坤, 孙未未

引用本文

黄坤, 孙未未. [基于缺失数据的交通速度预测算法](#)[J]. 计算机科学, 2024, 51(3): 72-80.

HUANG Kun, SUN Weiwei. [Traffic Speed Forecasting Algorithm Based on Missing Data](#)[J]. Computer Science, 2024, 51(3): 72-80.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[CheatKD:基于毒性神经元同化的知识蒸馏后门攻击方法](#)

CheatKD: Knowledge Distillation Backdoor Attack Method Based on Poisoned Neuronal Assimilation
计算机科学, 2024, 51(3): 351-359. <https://doi.org/10.11896/jsjcx.221200035>

[基于依赖类型剪枝的双特征自适应融合网络用于方面级情感分析](#)

Dual Feature Adaptive Fusion Network Based on Dependency Type Pruning for Aspect-based Sentiment Analysis

计算机科学, 2024, 51(3): 205-213. <https://doi.org/10.11896/jsjcx.230100035>

[异质信息网络中基于解耦图神经网络的社区搜索](#)

Community Search Based on Disentangled Graph Neural Network in Heterogeneous Information Networks

计算机科学, 2024, 51(3): 90-101. <https://doi.org/10.11896/jsjcx.221200029>

[基于启发式粗化算法的半监督图神经网络的训练加速框架及算法](#)

Framework and Algorithms for Accelerating Training of Semi-supervised Graph Neural Network Based on Heuristic Coarsening Algorithms

计算机科学, 2024, 51(3): 48-55. <https://doi.org/10.11896/jsjcx.221200158>

[基于对比图学习的跨文档虚假信息检测](#)

Contrastive Graph Learning for Cross-document Misinformation Detection

计算机科学, 2024, 51(3): 14-19. <https://doi.org/10.11896/jsjcx.230800063>

基于缺失数据的交通速度预测算法

黄坤 孙未未

复旦大学计算机科学技术学院 上海 200438

上海市数据科学重点实验室(复旦大学) 上海 200438

上海智能电子与系统研究院 上海 200438

(khuang20@fudan.edu.cn)

摘要 交通速度预测是智能交通系统的基础,可以缓解交通拥堵,节约公共资源,提高人们的生活质量。在真实情况下,采集到的交通速度数据通常存在缺失,而现有研究成果大多数只考虑了数据相对完整的场景。文章主要针对缺失场景下的交通速度数据进行研究,捕捉其中的时空相关性,并对未来交通速度进行预测。为了充分利用到交通数据的时空特征,提出了一种新的基于深度学习的交通速度预测模型。首先,提出了“还原-预测”算法,先使用自监督学习方法让模型还原缺失数据,再对交通速度进行预测;其次,引入了对比学习的方法,使得速度时间序列的特征表示更鲁棒;最后,模拟了不同数据缺失率的场景,通过实验验证了所提方法在各种缺失率下的预测准确率都优于现有方法,并设计了实验对对比学习方法和不同的还原算法进行分析,证明了所提方法的有效性。

关键词: 交通速度预测; 缺失数据还原; 图神经网络; 对比学习; 深度学习

中图分类号 TP301

Traffic Speed Forecasting Algorithm Based on Missing Data

HUANG Kun and SUN Weiwei

School of Computer Science, Fudan University, Shanghai 200438, China

Shanghai Key Laboratory of Data Science, Fudan University, Shanghai 200438, China

Shanghai Institute of Intelligent Electronics and Systems, Shanghai 200438, China

Abstract Traffic speed forecasting is the foundation of intelligent transportation system, which can ease traffic congestion, save public resources and improve people's quality of life. In real situations, the collected traffic speed data are usually missing, and most of the existing research results only consider the scenarios with relatively complete data. The paper focuses on the traffic speed data in the missing scenarios, captures the spatio-temporal correlation, and predicts the future traffic speed. In order to make full use of the spatio-temporal characteristics of traffic data, this study proposes a new deep learning-based traffic speed forecasting model. Firstly, a “recover-predict” algorithm is designed, which first uses a self-supervised learning method to enable the model to recover the missing data and then predict the traffic speed. Secondly, a contrastive learning method is introduced to make the feature representation of the speed time series more robust. Finally, the scenarios with different missing data rates are simulated, and experimental results show that the prediction accuracy of the proposed method outperform existing methods with various missing rates, and experiments are designed to analyze the comparative learning method and different recovery algorithms to prove the effectiveness of the proposed method.

Keywords Traffic speed forecasting, Recovery of missing data, Graph neural network, Contrastive learning, Deep learning

1 引言

随着城市化进程的加快,道路上的车辆数量日益增多,交通在人们日常生活中起着愈发重要的作用。交通速度评估是智能交通系统(ITS)^[1]的基础。高质量的交通速度预测也是常用的交通服务的基础,比如路线规划和地图导航。准确的

实时交通速度预测可以为出行者预警,缓解交通拥堵并降低人们的出行成本,节约公共资源,从而提高人们的生活质量。

交通速度预测问题具有很强的时空相关性,即一个路段的交通状况与其历史的观测值相关,相近路段的交通状况也会互相影响。图1展示了交通速度的时空相关性,为了直观起见,假设最左图是单向道的道路网络图,规定车辆从左往右

到稿日期:2023-01-09 返修日期:2023-04-30

基金项目:国家自然科学基金(62172107)

This work was supported by the National Natural Science Foundation of China(62172107).

通信作者:孙未未(wwsun@fudan.edu.cn)

行驶。其中时间维包括 $t \sim t+2$ 3 个时间片,空间维包括 8 个节点(以路段为节点,标号 1-8),节点之间的边表示车辆是否可以直接从一个路段行驶至另一个路段。每个节点颜色的深浅表示此刻所有行驶在该路段上的车辆的平均速度,颜色越深则车速越慢,即道路越拥堵。假设 t 时刻路段 1 先发生拥堵,车速变慢,随着时间的推移,和路段 1 关系较为密切的路段 3、路段 2 和路段 6 的拥堵程度明显上升,其他路段的车速也相应发生了变化。在这个例子中,一个节点的拥堵现象随着时间的发展通过路网影响到其他路段,这正是交通流的时空相关性的体现。然而,现实生活中道路的情况更为复杂。不仅需要考虑附近的节点,还需要考虑距离较远的节点,如公路路口和入口,而且路网中多个节点同时拥堵的情况也并不少见。这些拥堵节点的相互作用也应当被仔细地建模。如何从这些复杂、非线性的时空数据中提取时空关系是个极大的挑战。研究者们通常将交通速度预测问题视为一个图时空学习(Geometric Spatio-Temporal Learning)问题,并提出了一些方法^[2-6]来进行求解。

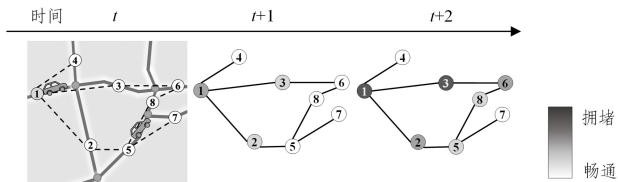


图 1 交通速度的时空相关性

Fig. 1 Spatio-Temporal correlation of traffic speed

然而,这些工作都只考虑了数据相对完整的场景,较少涉及针对有缺失数据的交通速度预测问题。目前的交通速度信息,主要通过道路上安装交通传感器(如环形线圈检测器)收集。这种方法的优点是,收集到的信息可以直接和路段绑定,预处理的开销很小。然而,传感器也存在着一定的局限性:一是容易出现硬件故障,导致数据缺失或产生数据噪声;二是由于安装成本比较高昂,因此只能在有限的位置上安装传感器。以上两个原因导致真实情况下的交通速度数据可能并不完整,基于完整速度数据的方法也不一定适配不完整的场景。

为了应对上述挑战,本文提出了一种基于深度学习的缺失数据场景下的交通速度预测模型。模型主要由两个子模型组成:还原子模型和预测子模型。其中,还原子模型负责还原缺失的数据,我们使用自监督学习方法对还原子模型进行训练,先将输入的数据进行随机掩码,并以掩码的部分作为监督信息让模型学习到对缺失数据的还原能力。预测子模型则利用还原好的数据对未来的交通速度进行预测。而考虑到原始数据本身可能存在噪声,同时还还原子模型也会引入一定的噪声,为了降低噪声对预测子模型的影响,我们使用对比学习^[7]的方法,利用交通速度具有时间周期性的特点,使时段相似的速度时间序列的特征表示更加接近。本文的主要贡献概括如下:

1)设计了“还原-预测”(Recover-Predict)算法,该算法包含还原子模型和预测子模型两个部分。先通过自监督学习方法让模型对缺失数据具有还原能力,再在还原数据的基础上

进行未来速度预测。具体地,将图信息加入原始的 Trans-former^[8]中作为还原子模型,使用 STGCN^[2]作为预测子模型。

2)引入对比学习的方法,让具有相似性质(如每星期的同一时间段)的速度时间序列的特征表示尽可能拉近,其他的特征表示尽可能远离,从而得到更鲁棒的速度时间序列特征表示。

3)在具有完整数据的数据集上生成了缺失场景下的数据集,模拟了不同的缺失率,并进行了实验,证明了所提方法在不同的缺失率下都具有可行性。

2 相关工作

交通速度预测在现实世界中是个非常重要的问题,很多研究人员都在这个问题上做过深入探索。

许多传统方法将交通速度预测视为时间序列预测问题,如差分整合移动平均自回归模型(ARIMA)^[9]和支持向量机(SVM)^[10],它们无法捕捉道路网络的复杂特征,也很难建立时空关联。随着近年来计算机硬件和大规模计算的发展,深度学习方法应用于交通预测领域成为可能^[11]。深度学习模型有更多的参数,相较于传统方法,其表达能力更强,准确性更高。Zang 等^[12]侧重于长期交通速度预测,并堆叠了 3 个 ConvLSTM^[13]层来提取多尺度时间特征。Chen 等^[14]建立了多尺度的网络模型,采用双向 GRU^[15]提取多尺度时间依赖性。而大部分交通状况指标是沿路网分布或者与固定的道路关联,可以将其自然地建模为图结构。使用深度图神经网络(GNN)^[16]进行交通速度预测是最近的研究趋势。GNN 忽略节点顺序,通过节点间的信息传递来建模图结构和图中的依赖关系。Kipf 等^[17]提出了图卷积网络(GCN),其借助谱图理论,将图像中的卷积操作引入图结构数据的处理。但是,GCN 需要将全图放在内存或显存中进行计算,无法处理大图,且扩展性较差。GraphSAGE^[18]在其基础上进行改进,将邻居节点的信息进行聚合得到每个节点的特征表示,在没有全图信息的情况下进行计算,可以高效地处理路网信息。Yu 等^[2]提出了时空图卷积神经网络(STGCN),首次使用纯卷积结构同时从图结构时间序列中提取时空特征。Graph WaveNet^[3]提出自适应邻接矩阵来捕获隐藏的空间依赖,并将图卷积和空洞因果卷积相结合。Liu 等^[4]设计了循环神经网络^[19](RNN)的变体 MSDR 来提取时间表示,并结合了注意力机制和 GNN 模块,提出了 GMSDR 预测交通状况。此外,还有一些工作从其他角度研究这个问题。Zheng 等^[5]提出的 GMAN 使用了多头注意力网络来模拟时空因素对交通状况的影响。Ji 等^[6]则从交通流的物理原理入手,建立了时空微分方程网络来进行预测。然而上述模型都没有考虑数据缺失的情况。

Liu 等的工作^[20]研究了数据缺失的场景,把路网交通速度序列视作具有缺失值的路段时间的二维矩阵。该工作使用非负稀疏矩阵分解(SNMF)^[21]来计算拓扑结构和节点属性之间的依赖关系,并在正则化过程中考虑了非线性时空相关性,以此来恢复缺失数据,最后使用 GraphSAGE 聚合图信息进行预测。但该工作没有充分利用时间序列的时序特征,也

无法很好地对路段特征进行表示。此外,还有一些工作对交通速度还原任务进行了研究。Chen 等^[22]考虑了更多因素间的依赖关系,以路段、时间和日期形成三维稀疏张量,并提出了最小化截断核范数(TNN)的低秩张量还原(LRTC)算法 LRTC-TNN。还有一些工作使用深度学习方法从历史数据中学习隐藏的依赖关系来生成缺失数据。Zhang 等^[23]设计了基于 LSTM 的序列到序列(Seq2Seq)^[24]模型 SSIM,有效补充了长度不定的缺失数据序列。Xu 等^[25]结合 GraphSage 和生成对抗网络(GAN)^[26]提出了 GA-GAN 模型,GraphSage 聚合时空信息,GAN 从聚合的信息中还原完整数据。

3 基本概念

定义 1(道路网络, Road Network, 简称路网) 道路网络可以建模为图 $G=(V, E, \mathbf{A})$ 。其中, V 为顶点集,表示路段的集合,集合的大小 $|V|=n$ 表示路段的数量; E 为边集,表示路网路段之间的连通性; $\mathbf{A} \in \{0, 1\}^{n \times n}$ 为图的邻接矩阵。

定义 2(路网拉普拉斯矩阵) 对于路网 G , 它的拉普拉斯矩阵定义为 $\mathbf{L} = \mathbf{D} - \mathbf{A}$, 其归一化形式为 $\mathbf{L} = \mathbf{I}_n - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$ 。其中 \mathbf{I}_n 为单位阵, $\mathbf{D} \in R^{n \times n}$ 是节点度数组成的对角矩阵。矩阵 \mathbf{L} 是半正定的。

定义 3(路网交通速度, Road Network Traffic Speed) 路段 i 在时间片 t 的交通速度为该时间片中通过该路段的所有车辆的速度的平均值, 记为 x_t^i 。则路网 G 在时间片 t 的

交通速度定义为 $\mathbf{X}_t = [x_t^1, x_t^2, \dots, x_t^n]^T \in R^{n \times 1}$ 。

定义 4(路网交通速度序列) 路网 G 连续的 n_h 个时间片 $(t-n_h+1) \sim t$ 的交通速度序列定义为 $\mathbf{X}_{t, n_h} = [\mathbf{X}_{t-n_h+1}; \dots; \mathbf{X}_t] \in R^{n \times n_h}$ 。

定义 5(缺失数据下的交通速度预测) 对于路网 G , 定义 $\mathbf{H}_{t, n_h} \in R^{n \times n_h}$ 为观测到的具有缺失数据的交通速度矩阵, $\mathbf{M}_{t, n_h} \in \{0, 1\}^{n \times n_h}$ 表示缺失值的位置, 其中 0 表示该值缺失。则 $\mathbf{H}_{t, n_h} = \mathbf{X}_{t, n_h} \odot \mathbf{M}_{t, n_h}$, 其中 \odot 表示矩阵点乘。缺失数据下的交通速度预测的定义为: 给定路网 G 和输入 \mathbf{H}_{t, n_h} , 求函数 F 来预测未来 n_p 个时间片的最大可能路网交通速度 \mathbf{Y}_{t+n_p, n_p} 。

$$F(\mathbf{H}_{t, n_h}) \approx \arg \max \log p(\mathbf{Y}_{t+n_p, n_p} | \mathbf{H}_{t, n_h}) \quad (1)$$

4 “还原-预测”算法

本章将详细介绍算法的模型结构和算法流程。算法的基本框架如图 2 所示, 模型主要包括还原子模型和预测子模型两个部分。算法先通过输入的缺失速度矩阵 \mathbf{H} 和路网图 G 构造正例 $\tilde{\mathbf{H}}$, 正例由两种方式构成: 以往星期同一时段的路网速度矩阵或噪声扰动后的自身速度矩阵。将原始输入和正例都通过还原子模型得到还原后的速度矩阵 \mathbf{H}_R 和 $\tilde{\mathbf{H}}_R$ 。预测子模型通过 \mathbf{H}_R 和 $\tilde{\mathbf{H}}_R$ 对未来时段的速度矩阵进行预测, 并通过对比学习对预测子模型输出的特征向量 \mathbf{v}^{k+1} 和 $\tilde{\mathbf{v}}^{k+1}$ 进行约束, 以得到更鲁棒的特征向量表示。

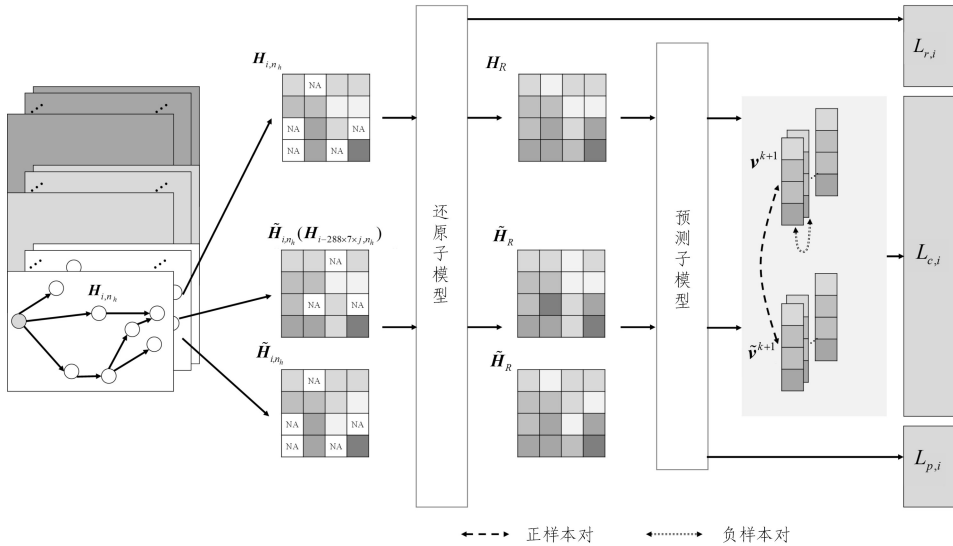


图 2 模型基本框架

Fig. 2 Basic framework of our model

下面将分别对还原子模型、预测子模型和对比学习 3 个部分进行详细介绍。

4.1 还原子模型

还原子模型通过自监督学习方法学习到对缺失数据的推理能力。对于给定缺失速度矩阵 \mathbf{H} 和缺失位置矩阵 \mathbf{M} , 我们随机产生一个新的掩码矩阵 $\mathbf{M}_H \in \{0, 1\}^{n \times n_h}$, 保证矩阵 \mathbf{M}_H 和 \mathbf{M} 中 0 的位置不相同。还原子模型以掩码后的缺失速度矩阵 $\mathbf{H}' = \mathbf{H} \odot \mathbf{M}_H$ 和路网 G 作为输入, 以掩码的信息 $\mathbf{H} \odot (1 - \mathbf{M}_H)$ 作为监督条件进行训练。训练好的还原子模型能够通过

已知的数据推理掩码的数据。而由于掩码是随机进行的, 且缺失数据和掩码数据来自同一个数据分布, 因此, 还原子模型对缺失数据也具备一定的还原能力。

如图 3 所示, 我们使用 Transformer^[8] 作为还原子模型的主要结构。Transformer 是序列到序列的模型, 其使用自注意力机制(Self-attention)来进行长距离的信息交换, 无视空间限制, 以获取更深层次的特征信息。它被广泛运用于自然语言处理、语音处理等领域中, 也通常被用来解决自监督任务。

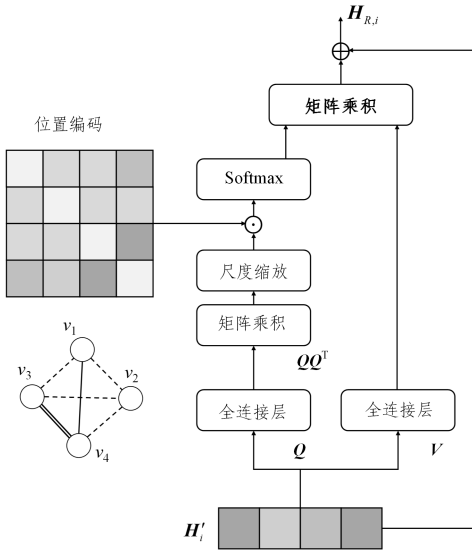


图3 还原子模型结构图

Fig. 3 Structure diagram of recover sub-model

自注意力机制是注意力机制的一种特殊形式,矩阵 Q, K, V 均来自同一输入加权变换。对于输入 $H' \in R^{n \times n_h}$, 有:

$$Attention(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_K}} \right) V \quad (2)$$

其中, $Q = H'W_Q, K = H'W_K, V = H'W_V$ 分别为 Query, Key 和 Value 矩阵; $W_Q, W_K, W_V \in R^{n_h \times d_K}$ 是可学习的权重矩阵, d_K 为输出的维度, 将点乘结果除以 $\sqrt{d_K}$ 以防止值过大。根据 Tsai 等^[27]的工作, Q 和 K 使用相同的矩阵, 可以有效减少模型的参数数量, 而实际性能变化却很小。

$$Attention(Q, V) = \text{softmax} \left(\frac{QQ^T}{\sqrt{d_K}} \right) V \in R^{n \times d_K} \quad (3)$$

我们也使用了多头注意力机制和残差连接, 通过不同的 W_Q 和 W_V 来生成不同的 Q 和 V :

$$MultiHead(H') = \text{Concat}(Attention(Q_1, V_1), \dots, Attention(Q_h, V_h))W^o \quad (4)$$

$$H' = H' + MultiHead(H') \quad (5)$$

其中, Concat 为向量连接操作, h 为注意力头的数量, $W^o \in R^{(h \times d_K) \times d_K}$ 为权重矩阵。

自注意力机制具有置换不变的性质。一个没有位置编码的 Transformer 可以被看作是一个处理完全图的 GNN, 每一个节点都聚合其他所有节点的信息。通常来说, 我们需要在 Transformer 中加入位置编码 (Position Encoding) 来标注每个元素在输入序列中的位置。然而, 图不是顺序结构, 节点间是多对多的关系, 并且位置的概念并不明确。因此, 我们使用节点结构相似性的信息 K_r 作为注意力分数的权重偏置项, 以此来引入图的位置编码。

$$PosAtt(Q, V, K_r) = \text{normalize} \left(\exp \left(\frac{QQ^T}{\sqrt{d_K}} \right) \odot K_r \right) V \quad (6)$$

其中, \exp 是以自然常数 e 为底的指数函数, normalize 为 L1 归一化函数。本文使用 p 步随机游走核和全连接核线性加权的方式来刻画节点的位置编码, 即

$$K_r = k_1 \times K_{RW} + k_2 \times K_{FC} = k_1 \times (I - \gamma L)^p + k_2 \times J_n \quad (7)$$

其中, k_1 和 k_2 为超参数, 表示两种核的权重, γ 为随机游走核的参数, J_n 表示 n 阶全 1 方阵。

对于节点 i , 我们有:

$$PosAtt(Q, V, K_r)_i = \sum_{j=1}^n \frac{\exp(QQ^T / \sqrt{d_K}) \times K_r(i, j)}{\sum_{j'=1}^n \exp(QQ^T / \sqrt{d_K}) \times K_r(i, j')} V_j \quad (8)$$

在前馈的过程中, 我们使用归一化系数 $D^{-\frac{1}{2}}$ 来减少高联通度的联通分量的影响。

$$H' = H' + D^{-\frac{1}{2}} PosAtt(Q, V, K_r) \quad (9)$$

还原子模型的输出为还原了缺失数据的速度矩阵, 记为 H_R 。模型的损失函数 L_r 为:

$$L_r = \frac{1}{P} \sum_{(i,j), M_H(i,j)=0} |H_R(i, j) - H(i, j)| \quad (10)$$

其中, P 为掩码位置的数量。

4.2 预测子模型

预测子模型以还原子模型输出的 H_R 和路网图 G 作为输入, 对未来的路网交通速度序列进行预测。我们使用 STGCN^[2] 作为预测子模型的主要结构。STGCN 是时空图卷积神经网络模型, 其从图结构时间序列中提取时空特征, 并使用纯卷积结构, 参数少, 可以高效处理大规模图。

STGCN 主要由时域门控卷积和空域图卷积组成, 分别用来提取时间特征和空间特征。其结构示意图如图 4 所示。

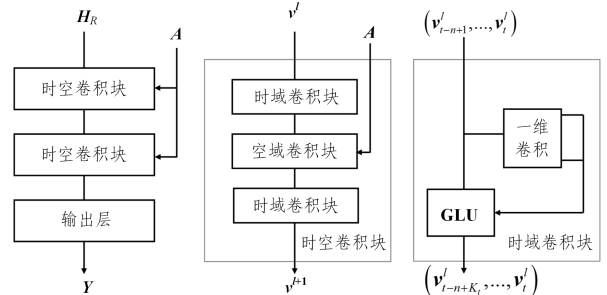


图4 预测子模型结构图

Fig. 4 Structure diagram of predictive submodel

4.2.1 时间特征提取

和 CNN 类似, 我们把输入 $H_R \in R^{n_h \times n \times C_{in}}$ 的每个时间片每个节点上的每一维特征视为一个通道。初始时, 只包含速度这一个特征, 故通道数为 1 ($C_{in} = 1$)。我们对每 K_t 个相邻的时间片使用卷积核 $\Gamma \in R^{K_t \times C_{in} \times (2C_{out})}$ 来捕捉交通流的动态时间特征, 其中 $2C_{out}$ 为输出的通道数。由于使用的是没有填充 (Padding) 的卷积方式, 因此每次卷积会将输入序列长度减少 $K_t - 1$ 。

为了得到更深层次的特征, 我们使用一个门控线性单元对特征进行组合。 H_R 通过卷积核 Γ 得到大小为 $(n_h - K_t + 1) \times n \times (2C_{out})$ 的特征矩阵, 将其按照通道顺序切分成大小均为 $(n_h - K_t + 1) \times n \times C_{out}$ 的特征矩阵 P 和 Q , 则

$$\Gamma H_R = P \odot \sigma(Q) \in R^{(n_h - K_t + 1) \times n \times C_{out}} \quad (11)$$

其中, \odot 表示矩阵点乘, σ 表示激活函数, 本文使用 sigmoid 函数。相比常用的基于 RNN 的时间特征提取结构, CNN 速度更快, 结构更简单, 对序列动态变化的响应更迅速, 更适用于

图结构的时间序列。

4.2.2 空间特征提取

基于谱图理论,我们引入图核函数 Θ 和图卷积算子 $\ast G$ 对图信号 \mathbf{h} (即某一时间片的路网交通速度向量) 进行运算,并使用切比雪夫多项式对其近似展开,

$$\Theta_{\ast G} \mathbf{h} = \Theta(\mathbf{L}) \mathbf{h} \approx \sum_{k=0}^{K-1} \theta_k T_k(\tilde{\mathbf{L}}) \mathbf{h} \quad (12)$$

其中, K 为展开的阶数, $\theta_k \in \mathbb{R}^K$ 为多项式系数, $\tilde{\mathbf{L}} = \frac{2\mathbf{L}}{\lambda_{\max}} - \mathbf{I}_n$ 为缩放后的路网拉普拉斯矩阵, λ_{\max} 为矩阵的最大特征值。 $T_k(x)$ 为 k 阶切比雪夫多项式,其递归定义为 $T_0(x) = 1$, $T_1(x) = x$, $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$ 。切比雪夫多项式相当于使用卷积核对图中每个节点提取了其 $0 \sim K-1$ 阶邻居的特征,相比直接变换到谱域上进行图卷积,这种方式不用将矩阵特征值分解,计算效率较高。

为了进一步加速计算,我们取 $K=2$,并假设 $\lambda_{\max}=2$,则

$$\Theta_{\ast G} \mathbf{h} \approx \theta_0 \mathbf{h} - \theta_1 (\mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}) \mathbf{h} \quad (13)$$

将式 $\theta = \theta_0 = -\theta_1$ 和式 $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_n$, $\tilde{\mathbf{D}}_{ij} = \sum_j \tilde{\mathbf{A}}_{ij}$ 代入,图卷积运算最终化简成:

$$\Theta_{\ast G} \mathbf{h} \approx \theta (\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}}) \mathbf{h} \quad (14)$$

使用时空卷积块来实现尺度缩放和特征提取,在每个卷积块中使用层归一化以防止过拟合。时空卷积块由两个时域卷积块和夹在它们中间的一个空域卷积块组成,假设第 l 个卷积块的输入为 $\mathbf{v}^l \in \mathbb{R}^{n_b \times n \times c^l}$,输出为 $\mathbf{v}^{l+1} \in \mathbb{R}^{(n_b - 2(K_l - 1)) \times n \times c^{l+1}}$,则

$$\mathbf{v}^{l+1} = \Gamma_1^l \text{ReLU}(\Theta_{\ast G}(\Gamma_0^l \mathbf{v}^l)) \quad (15)$$

其中, Γ_0^l 和 Γ_1^l 分别是第 l 块的两个时域卷积块, $\Theta_{\ast G}$ 是空域卷积块。整个 STGCN 有 k 个时空卷积块,定义第 k 个时空卷积块的输出 \mathbf{v}^{k+1} 为模型的时空特征表示,使用一个线性层将输出映射为 $\mathbf{Y} \in \mathbb{R}^{n_p \times n}$,并将其作为预测的未来时间片的速度矩阵。定义 L_p 为预测子模型的损失函数,则

$$L_p = \frac{1}{|D_s|} \sum_{(\mathbf{u}, \hat{\mathbf{v}}) \in D_s} |\mathbf{Y} - \hat{\mathbf{Y}}| \quad (16)$$

其中, D_s 为数据集, $\hat{\mathbf{Y}}$ 是真实的路网速度矩阵。

4.3 融合对比学习的损失函数

交通数据有着天然的时空相关性。比如,每周工作日早高峰(7-9点)和晚高峰(17-19点)车流量很大,早高峰车辆聚集在从住宅区到工业区的路段上,而晚高峰则相反。周末也有着类似的规律,在从住宅区到商业区的路段上会有更多的车辆聚集。

图5所示为传感器信号站1在连续两周内的交通速度变化折线图,其中纵轴表示交通速度,横轴表示连续时间片的编号(时间片以5min为单位划分,一天包含288个时间片,一周包含2016个时间片)。从图中可以看出,周一到周五的路段交通速度变化具有相似性,但周末和工作日的交通速度变化则有很大差别。图6为传感器信号站1-8在连续6周内同一时间段的交通速度变化曲线,其中纵轴表示交通速度,横轴表示周编号。从图中可以看出,对于不同周的同一时间段,各个路段的交通速度变化都比较平滑。这说明,路段交通速度的变化同时具有日周期性和周周期性,但日周期性和是否为工作日的关联极大。因此,我们选择利用周周期性进行对比学习。

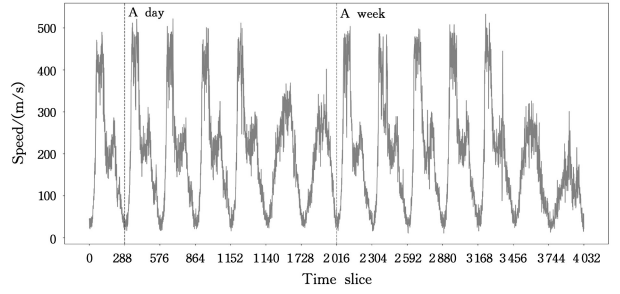


图5 单传感器信号站两周内交通速度变化

Fig. 5 Two-week traffic speed change at a single sensor station

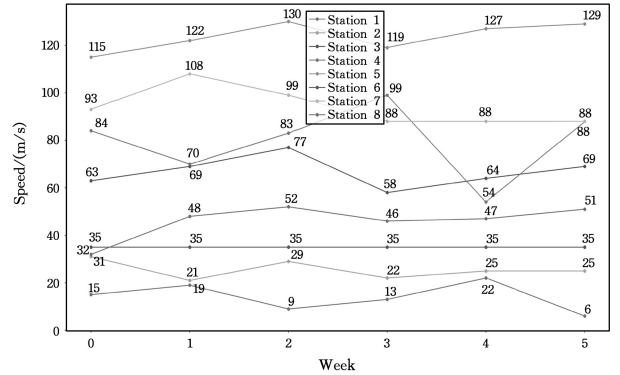


图6 多传感器信号站交通速度周变化

Fig. 6 Weekly traffic speed change at multi-sensor station

我们希望,在不同周周期内,同一路段对应时段的速度向量的特征表示尽可能接近,不同时段特征表示尽可能远离。本文使用对比学习的方法来优化我们的特征表示。对比学习是一种自监督学习方式,通过构造正例和负例来进行训练。本文使用的两种构造正例的方式如下:

1) 使用以往星期同一路段同一时段的速度向量作为正例。

2) 对当前的输入加入噪声扰动以生成正例。对于当前输入速度矩阵 \mathbf{H} ,随机在矩阵的某些位置加入服从正态分布的噪声,使得 $\tilde{\mathbf{H}}_{i,j} = \mathbf{H}_{i,j} + \mathbf{N}(0, 0.1)$ 。

以上两种方式的正例的占比为7:3。负例为同一训练批次(Batch)内除本身外的其他所有样本。使用 InfoNCE^[28] 作为对比学习损失,通过计算正负样本在预测子模型最后一层的特征表示 \mathbf{v}^{k+1} 的相似度来拉近正样本、拉远负样本。

$$L_c = -\log \frac{\exp(\mathbf{v}^{k+1} \cdot \mathbf{v}_+ / \tau)}{\sum_{i=1}^B \exp(\mathbf{v}^{k+1} \cdot \mathbf{v}_i / \tau)} \quad (17)$$

其中, \mathbf{v}_+ 是正例的特征表示; \mathbf{v}_i 是负例的特征表示; B 为负例的数量; τ 是温度超参数,用于控制模型对负样本的区分度。

综上所述,整个模型的总损失函数为还原损失、预测损失和对比学习损失的线性加权和,具体的计算式如下:

$$L = L_r + \alpha \times L_p + \beta \times L_c \quad (18)$$

其中, α 和 β 是权重超参数。

5 实验

5.1 数据集介绍

本文基于真实高速公路数据集 PeMSD4 来验证所提模型。

PeMSD4 是由部署在美国加利福尼亚州旧金山湾区的 3848 个传感器所收集的实时高速公路交通数据,时间跨度为 2018 年 1 月至 2 月,涉及 29 条道路。我们忽视了一些冗余传感器,使每个相邻传感器之间的距离都超过 3.5 英里,并根据传感器的位置分布将道路重新定义分割成了 307 个路段。图 7 所示为传感器站的地理位置分布。我们将原始采样频率为 30 秒/次的的数据以每 5 min 为一个时间片进行汇总,并以 12 个时间片为单位进行预测。即用上一小时内每 5 min 所有路段的速度矩阵作为输入,预测下一小时内每 5 min 所有路段的速度。

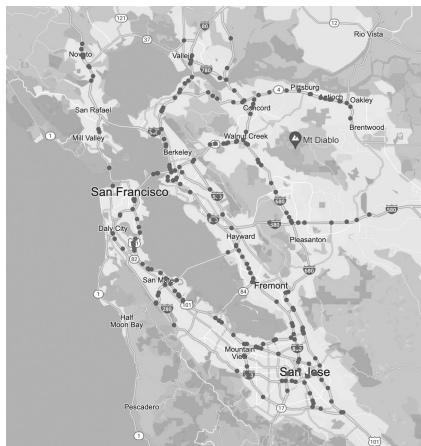


图 7 PeMSD4 传感器网络

Fig. 7 Sensor network of PeMSD4

为了建模稀疏情况下的速度矩阵,基于 PeMSD4 数据集,我们对输入的道路速度矩阵中 20% 的数据进行随机掩码(需要预测的真实输出不进行掩码),得到稀疏情况下的道路速度数据集 PeMSD4₂₀。相似地,PeMSD4₃₀ 和 PeMSD4₄₀ 分别表示随机掩码掉 30% 和 40% 的输入速度矩阵信息后得到的数据集。

本文按照时间顺序以 7:1:2 的比例将数据集划分成为训练集、验证集和测试集,其分别包含 11 877、1 696 和 3 392 条数据样本。并且使用训练集上交通速度的均值和方差对 3 个数据集进行归一化处理。

5.2 评价指标和对比方法

本文使用 3 个回归任务常用的指标来客观地评估模型的性能,分别是平均绝对误差(Mean Absolute Error, MAE)、平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)和均方根误差(Root Mean Squared Error, RMSE)。它们的具体计算公式分别为:

$$MAE = \frac{1}{m} \sum_{k=1}^K |Y_k - \hat{Y}_k| \quad (19)$$

$$MAPE = \frac{1}{m} \sum_{k=1}^K \frac{|Y_k - \hat{Y}_k|}{|Y_k|} \times 100\% \quad (20)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{k=1}^K (Y_k - \hat{Y}_k)^2} \quad (21)$$

在计算上述指标时,对于每个样本,我们忽略表示值不存在的 0 元素或 INF 元素,再求均值。

为了验证本文模型的有效性,我们选取了以下 8 种对比方法进行对比实验:

1) ARIMA^[9]: 差分整合移动平均自回归模型,对具有平稳性的时间序列的预测较为有效。在本文中,为了保证时序数据的平稳性,先使用滑动平均将缺失值还原,再进行预测。

2) SVR^[10]: 支持向量回归模型。本文对需要进行预测的每个时间片都独立训练一个 SVR 模型,并将预测结果按顺序拼接起来得到一段连续时间的速度预测。

3) GRU^[15]: 门控循环单元网络。本文使用单层的 GRU 网络和单层的全连接层进行预测。

4) STGCN^[2]: 由时域卷积块和空域卷积块两种结构组合而成的一种时空图卷积神经网络。

5) GWNet^[3]: 使用空洞因果卷积,以一定的步长在时间序列上跳跃卷积,来捕捉长距离序列的依赖关系;同时使用图卷积网络来捕捉数据的空间关系。

6) GMAN^[5]: 使用时空注意力模块组成的“编码-解码”结构的神经网络模型,并使用门控融合机制自适应地融合时间和空间表示。

7) GMSDR^[4]: 一种全新的递归神经网络变体,在图上使用了多步依赖关系。网络中显式地将多个历史时间步作为输入来进行预测。

8) SNMF-GraphSAGE^[20]: 对有缺失元素的速度矩阵使用非负矩阵分解进行恢复,然后使用 GraphSAGE 模型进行预测。

在上述 8 种对比方法中,ARIMA 和 SVR 是传统的机器学习时间序列算法;GRU 是对时间序列进行预测的经典深度学习学习方法,但是其没有利用到路网信息。模型 STGCN, GWNet, GMAN 和 GMSDR 融合了时间和空间信息,是目前效果较好的、解决完整数据场景下的交通流预测问题的算法。SNMF-GraphSAGE 则考虑了缺失数据的场景,先进行数据恢复后再预测。

此外,我们还选取了如下 3 种缺失数据还原算法来与所提方法的还原性能进行对比:

1) LRTC-TNN^[22]: 一种在低秩张量框架上提出的时空交通数据还原算法。算法设计了一种新的截断核范数来更好地描述时空交通数据中的隐藏模式。

2) SSIM^[23]: 一种基于 Seq2Seq 结构的深度学习模型,通过 LSTM 网络来建模时序信息。同时,SSIM 提出了可变长度滑动窗口算法来大量生成训练样本。

3) GA-GAN^[25]: 一种结合 GraphSAGE 和 GAN 的缺失交通数据还原算法。先通过路段历史数据来重建路网,然后使用 GraphSAGE 提取路段的时空信息,最后利用 GAN 进行还原。

在上面提及的 3 种方法中,LRTC-TNN 和 GA-GAN 是针对时空交通数据提出的还原算法,SSIM 是通用的时间序列还原算法。同时,我们将这 3 种方法作为还原子模型嵌入本文提出的“还原-预测”算法中,并以 STGCN 作为预测子模型,形成了 3 种新的对比方法 LRTC-TNN-STGCN, SSIM-STGCN 和 GA-GAN-STGCN。

5.3 实验设置

在还原子模型中,掩码矩阵 M_H 中 0 的数量为矩阵大小的 20%,即以 20% 的数据作为自监督信息来训练还原子

模型;核权重参数 $k_1=1, k_2=0.5$,核参数 $p=1, \gamma=0.5$ 。在预测子模型中,设置输出通道数 $C_{out}=128$,并堆叠 $k=2$ 个时空卷积块。在损失函数中,取温度系数 $\tau=0.07$,损失函数的权重 $\alpha=1, \beta=10$ 。

在训练过程中,首先仅训练还原子模型,使还原子模型能为预测子模型提供较高质量的还原结果,然后再同时优化还原损失和对比学习损失。

5.4 实验及结果分析

5.4.1 基于缺失数据的交通速度预测性能实验

表 1 列出了不同模型在该任务上的性能比较结果,可以看出,本文模型在大部分指标上都具有更小的预测误差。例如在 PeMSD₄₂₀ 数据集上,本文模型在平均绝对误差上比效果次优的 GWNet 模型降低了 0.385,在平均绝对百分比误差和均方根误差上分别降低了 0.465% 和 0.704。本文模型的优势主要体现在以下几个方面:

1) 相比 ARIMA 和 SVR 等传统的统计方法,或 GRU 等

只用到了序列信息的方法,本文方法更有效地利用了时空信息,效果更好。

2) 相比 STGCN 等融合了时空信息的神经网络,本文方法显式地将缺失数据还原作为一个任务拆分出来,并使用一个还原子模型去解决它。这种方式在具有更好的可解释性的同时,也能使模型更好地捕捉到输入信息之间的时空关系。

3) 相比 SNMF 和 LRTC-TNN 这类无监督的还原方法,所提还原方法利用自身的速度信息作为监督,更能捕捉到时空数据的内在关联。相比 SSIM 这种通用的时间序列还原算法,本文在还原子模型中加入了路网信息,使其能够结合周围路段的交通速度数据来进行还原。相比 GA-GAN 这种结合时空信息的还原方法,所提自监督还原方法能充分地利用样本数据,生成大量不同缺失位置的输入样本来进行训练。因而,本文方法具有更好的还原效果,能够为后续的预测子模型提供更准确的信息。

表 1 不同方法的性能比较

Table 1 Performance comparison of different methods

Method	PeMSD ₄₂₀			PeMSD ₄₃₀			PeMSD ₄₄₀		
	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE
ARIMA	33.089	21.705	49.116	33.192	23.968	49.653	33.135	24.710	49.469
SVR	30.933	22.084	46.937	31.320	22.665	47.755	31.997	23.091	48.534
GRU	27.057	19.312	41.024	27.641	20.484	41.765	28.439	21.211	42.882
STGCN	19.962	14.250	30.872	20.226	13.937	31.287	20.057	14.560	31.273
GWNet	19.624	14.180	31.003	19.730	14.025	30.676	20.049	14.767	31.435
GMAN	19.738	14.694	29.398	19.884	14.835	29.595	19.990	14.980	29.695
GMSDR	19.835	14.377	30.049	20.040	14.708	31.616	20.229	15.419	30.622
SNMF-Graph SAGE	20.270	14.198	31.803	20.669	14.356	32.511	20.857	14.863	32.919
Ours	19.239	13.715	29.299	19.298	13.698	29.379	19.623	14.519	29.690
LRTC-TNN-STGCN	19.510	13.797	30.593	19.616	13.957	30.640	20.813	14.963	30.512
SSIM-STGCN	19.688	13.977	29.852	19.608	14.035	29.889	20.531	14.727	30.977
GA-GAN-STGCN	19.936	13.834	31.540	20.156	13.884	31.693	21.195	14.740	32.915

表 1 的数据也说明了所提模型在不同缺失率的数据下都具有较好的效果。可以看到,当数据的缺失率升高时,数据包含的可用信息量减少,模型的效果也会变差。表 1 中的数据表明,缺失率从 20% 上升到 30% 时效果变化并不明显,而从 30% 上升到 40% 时效果变化较大。这一方面是因为当缺失率在 20%~30% 时,输入仍能为模型提供较为充足的信息;另一方面是因为当缺失率过高时,还原子模型的误差将会

累计到预测子模型上,导致最终的预测结果较差。

5.4.2 缺失数据还原性能实验

表 2 比较了各种还原方法和本文方法的还原性能。结果说明通过自监督掩码训练的还原子模型对缺失数据具有还原能力。同时,本文方法结合了路网信息,且更充分地利用了未缺失的数据作为监督信号,能更好地捕捉数据间的时空相关性,在各项指标上都有较大的优势。

表 2 不同方法的还原性能比较

Table 2 Recovery performance comparison of different methods

Method	PeMSD ₄₂₀			PeMSD ₄₃₀			PeMSD ₄₄₀		
	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE
SNMF	27.484	25.112	37.279	27.691	25.249	37.523	31.623	30.464	42.579
LRTC-TNN	21.207	19.163	31.884	24.346	19.087	35.423	28.067	36.095	38.208
SSIM	23.122	22.645	34.172	25.347	21.049	38.299	29.256	28.393	39.117
GA-GAN	24.758	23.071	36.589	26.501	23.587	38.166	28.250	28.467	40.519
Ours	18.673	16.981	27.131	18.705	16.018	27.346	21.392	20.040	31.083

5.4.3 “还原-预测”算法普适性实验

为了证明“还原-预测”算法的普适性,我们将所提方法中的 STGCN 替换成了不同的预测模型进行了实验。如表 3 所列,我们以数据集 PeMSD₄₂₀ 为例进行了测试,其中 RP@GWNet, RP@GMAN 和 RP@GMSDR 分别表示使用模型

GWNet, GMAN 和 GMSDR 作为预测子模型时的实验效果。相较于原始的预测方法,结合本文提出的还原子模型后效果都有一定程度的提升。以 MAE 指标为例, GWNet, GMAN, GMSDR 和 STGCN 结合还原子模型后比原先的表现分别提升了 1.10%, 0.51%, 0.83% 和 3.62%。还原子模型利用了

大量的自监督数据对数据之间的依赖关系进行建模,从而能给后续的预测子模型提供更优质的输入数据。实验结果证明结合还原子模型能更好地捕捉到输入信息之间的时空关系。

表3 不同预测模型的性能

Table 3 Performance of different prediction models

Method	PeMSD ₄₂₀		
	MAE	MAPE/%	RMSE
GWNet	19.624	14.180	31.003
GMAN	19.738	14.694	29.398
GMSDR	19.835	14.377	30.049
STGCN	19.962	14.250	30.872
RP@GWNet	19.409	13.784	30.324
RP@GMAN	19.638	15.091	29.360
RP@GMSDR	19.671	13.995	29.562
Ours(RP@STGCN)	19.239	13.715	29.299

STGCN在数据集 PeMSD₄₂₀上具有最好的效果,RP@STGCN相比 STGCN在 MAE,MAPE/%,RMSE这3项指标上分别提升了3.62%,3.75%,5.10%。这说明在近年来表现较好的时空预测模型中,STGCN是最适合本文算法的模型。因此,本文将作为主要的预测子模型,在其他缺失率的数据集上进行了更深入的实验。

5.4.4 对比学习有效性实验

表4列出了对比学习损失的消融实验的结果,其中方法Ours*为去除对比学习损失的“还原-预测”模型,表中指标为平均绝对误差。对比学习损失分别在 PeMSD₄₂₀,PeMSD₄₃₀和 PeMSD₄₄₀数据集上使模型的MAE提升了1.17%,1.10%和1.00%。在数据缺失率提高的数据集上加入对比学习损失后的提升效果呈下降趋势的原因主要在于,随着缺失率的增大,可用于对比学习的数据相应减少。与没有加入对比学习损失的模型相比,对比学习损失的加入,在不同缺失率的数据集上都可以提高特征表示的鲁棒性,从而提升模型的预测准确度。

表4 对比学习损失的消融实验

Table 4 Ablation study of contrastive learning loss

Method	PeMSD ₄₂₀	PeMSD ₄₃₀	PeMSD ₄₄₀
	MAE	MAE	MAE
Ours	19.239	19.298	19.623
Ours*	19.467	19.512	19.821

结束语 本文从缺失数据场景的角度出发,提出了一种基于深度学习的交通速度预测算法。算法利用自监督学习方法并融合对比学习方法,使模型能挖掘出道路速度数据中更深的内在联系,从而解决数据缺失的问题。本文方法显式地将缺失数据还原作为一个任务拆分出来,并使用一个还原子模型来解决它。相比一般的解决交通速度预测问题的方法,本文方法对缺失数据更具有针对性;相比无监督的还原方法,本文方法具有更好的还原性能。最终在多种缺失率的数据集上证明了本文方法具有更好、更全面的性能。在未来的研究工作中,我们将主要聚焦如下两个方面:一是在更多的数据集上测试模型的有效性;二是设计与“还原-预测”算法更加适配的还原、预测子模型结构,以进一步提升预测性能。

参考文献

[1] ZHANG J, WANG F, WANG K, et al. Data-Driven Intelligent

Transportation Systems: A Survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(4): 1624-1639.

- [2] YU B, YIN H, ZHU Z. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting[C]// Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence(IJCAI 2018). Stockholm: Morgan Kaufmann, 2018: 3634-3640.
- [3] WU Z, PAN S, LONG G, et al. Graph WaveNet for Deep Spatial-Temporal Graph Modeling[C]// Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence(IJCAI 2019). Macao: Morgan Kaufmann, 2019: 1907-1913.
- [4] LIU D, WANG J, SHANG S, et al. MSDR: Multi-Step Dependency Relation Networks for Spatial Temporal Forecasting[C]// Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Washington: ACM, 2022: 1042-1050.
- [5] ZHENG C, FAN X, WANG C, et al. GMAN: A Graph Multi-Attention Network for Traffic Prediction[C]// Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence(AAAI 2020). New York: AAAI, 2020: 1234-1241.
- [6] JI J, WANG J, JIANG Z, et al. STDEN: Towards Physics-guided Neural Networks for Traffic Flow Prediction[C]// Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence(AAAI 2022). AAAI, 2022: 4048-4056.
- [7] HADESELL R, CHOPRA S, LECUN Y. Dimensionality Reduction by Learning an Invariant Mapping[C]// 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR 2006). New York: IEEE, 2006: 1735-1742.
- [8] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All You Need[C]// Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017. Long Beach: MIT Press, 2017: 5998-6008.
- [9] YU G, ZHANG C. Switching ARIMA Model Based Forecasting for Traffic Flow[C]// 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing. Montreal: IEEE, 2004: 429-432.
- [10] JEONG Y S, BYON Y J, CASTRO-NETO M M, et al. Supervised Weighting-Online Learning Algorithm for Short-Term Traffic Flow Prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(4): 1700-1707.
- [11] LV Y, DUAN Y, KANG W, et al. Traffic Flow Prediction With Big Data: A Deep Learning Approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(2): 865-873.
- [12] ZANG D, LING J, WEI Z, et al. Long-Term Traffic Speed Prediction Based on Multiscale Spatio-Temporal Feature Learning Network[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(10): 3700-3709.
- [13] SHI X, CHEN Z, WANG H, et al. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting[C]// Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015. Montreal: MIT Press, 2015: 802-810.
- [14] CHEN D, YAN X, LIU X, et al. A Multiscale-Grid-Based Stacked Bidirectional GRU Neural Network Model for Predic-

- ting Traffic Speeds of Urban Expressways[J]. IEEE Access, 2021,9:1321-1337.
- [15] CHUNG J, GULCEHRE C, CHO K, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling [C]//NIPS 2014 Workshop on Deep Learning, 2014.
- [16] GORI M, MONFARDINI G, SCARSELLI F. A New Model for Learning in Graph Domains[C]//Proceedings of 2005 IEEE International Joint Conference on Neural Networks. Montreal: IEEE, 2005, 2:729-734.
- [17] KIPF T N, WELING M. Semi-Supervised Classification with Graph Convolutional Networks[C]//Conference Track Proceedings of the 5th International Conference on Learning Representations (ICLR 2017). Toulon, 2017.
- [18] HAMILTON W L, YING Z, LESKOVEC J. Inductive Representation Learning on Large Graphs[C]//Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017. Long Beach: MIT Press, 2017:1024-1034.
- [19] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [20] LIU J, ONG G P, CHEN X. GraphSAGE-Based Traffic Speed Forecasting for Segment Network With Sparse Data[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(3):1755-1766.
- [21] HUANG K, SIDIROPOULOS N D, SWAMI A. Non-Negative Matrix Factorization Revisited: Uniqueness and Algorithm for Symmetric Decomposition[J]. IEEE Transactions on Signal Processing, 2013, 62(1):211-224.
- [22] CHEN X, YANG J, SUN L. A Nonconvex Low-Rank Tensor Completion Model for Spatiotemporal Traffic Data Imputation [J]. Transportation Research Part C: Emerging Technologies, 2020, 117:102673.
- [23] ZHANG Y, THORBURN P J, XIANG W, et al. SSIM—A Deep Learning Approach for Recovering Missing Time Series Sensor Data[J]. IEEE Internet Things Journal, 2019, 6(4):6618-6628.
- [24] SUTSKEVER I, VINYALS O, LE Q V. Sequence to Sequence Learning with Neural Networks[C]//Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014. Montreal: MIT Press, 2014:3104-3112.
- [25] XU D, PENG H, WEI C, et al. Traffic State Data Imputation: An Efficient Generating Method Based on the Graph Aggregator [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(8):13084-13093.
- [26] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative Adversarial Nets[C]//Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014. Montreal: MIT Press, 2014:2672-2680.
- [27] TSAI Y H H, BAI S, YAMADA M, et al. Transformer Dissection: A Unified Understanding for Transformer's Attention via the Lens of Kernel[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: ACL, 2019:4343-4352.
- [28] HE K, FAN H, WU Y, et al. Momentum Contrast for Unsupervised Visual Representation Learning [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020). Seattle: IEEE, 2020:9729-9738.



HUANG Kun, born in 1997, postgraduate, is a student member of CCF (No. O9885G). Her main research interests include spatial-temporal data mining and so on.



SUN Weiwei, born in 1973, Ph.D. professor, is a senior member of CCF (No. 08792S). His main research interests include big spatial-temporal data and so on.

(责任编辑:何杨)