

STCTN: 一种基于时域偏倚校正与空域因果传递的时空因果表示学习方法

邓攀¹⁾ 刘俊廷¹⁾ 王晓³⁾ 贾晓丰⁴⁾ 赵宇²⁾
汪慕澜²⁾ 戴星原⁵⁾

¹⁾(北京航空航天大学软件学院 北京 100191)

²⁾(北京航空航天大学计算机学院 北京 100191)

³⁾(安徽大学人工智能学院 合肥 230601)

⁴⁾(北京市大数据中心数据管理部 北京 100101)

⁵⁾(中国科学院自动化研究所复杂系统管理与控制国家重点实验室 北京 100190)

摘要 从时空数据中有效地提取特征表示是时空数据挖掘的基础. 现有时空表示学习方法过于关注时空序列中的统计相关性, 易受虚假相关性的影响, 难以提取无偏、鲁棒的特征表示. 本文基于结构因果模型建模时空数据的生成过程, 分析观测数据中虚假相关性的成因, 提出了一种基于时域偏倚校正与空域因果传递的时空因果表示学习方法. 首先基于后门调整消除时域的虚假相关性, 然后构建因果传递网络消除空域的虚假相关性, 最后利用下游特征解码器将因果特征表示应用于下游任务中. 在两个真实数据集上的实验表明, 本文所提时空表示学习方法有效避免了虚假相关性的干扰, 增强了模型的稳定性, 使其在两个下游预测任务中对数据稀疏节点的预测误差分别降低了3%和10%.

关键词 时空表示学习; 结构因果模型; 虚假相关性; 后门调整; 因果关系

中图法分类号 TP391 **DOI号** 10.11897/SP.J.1016.2023.02535

STCTN: A Spatio-temporal Causal Representation Learning Method Based on Temporal Bias Adjustment and Spatial Causal Transition

DENG Pan¹⁾ LIU Jun-Ting¹⁾ WANG Xiao³⁾ JIA Xiao-Feng⁴⁾ ZHAO Yu²⁾
WANG Mu-Lan²⁾ DAI Xing-Yuan⁵⁾

¹⁾(School of Software, Beihang University, Beijing 100191)

²⁾(School of Computer Science and Engineering, Beihang University, Beijing 100191)

³⁾(School of Artificial Intelligence, Anhui University, Hefei 230601)

⁴⁾(Data Management Department, Beijing Big Data Centre, Beijing 100191)

⁵⁾(The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100191)

Abstract Spatio-temporal data, sampled from complex dynamical systems, is ubiquitous in real world, e. g., traffic flow, meteorological records and energy consumption. Learning effective feature representation from spatiotemporal data is the foundation of spatiotemporal data mining.

收稿日期: 2022-11-06; 在线发布日期: 2023-07-13. 邓攀, 博士, 副研究员, 主要研究领域为时空数据挖掘、交通流预测、因果性学习. E-mail: pandeng@buaa.edu.cn. 刘俊廷, 博士研究生, 主要研究领域为时空数据挖掘、交通流预测、因果性学习. 王晓, 博士, 教授, 主要研究领域为社会计算、平行车联网、认知自动驾驶. 贾晓丰(通信作者), 博士, 高级工程师, 主要研究领域为复杂系统优化、时空数据挖掘. E-mail: jiaxf@xj.beijing.gov.cn. 赵宇, 博士研究生, 主要研究领域为时空数据挖掘、交通流预测、因果性学习. 汪慕澜, 硕士研究生, 主要研究领域为时空数据挖掘、交通流预测、因果性学习. 戴星原, 博士, 助理研究员, 主要研究领域为人工智能、智能交通系统、认知自动驾驶.

The existing models overemphasize the statistical correlations in the spatiotemporal data and are susceptible to spurious correlations, which makes it hard to extract unbiased and robust feature representation. We model the generation process of spatiotemporal data based on the structural causal model (SCM), analyze the causes of spurious correlations in observations, and propose a spatiotemporal causal representation learning method based on temporal bias adjustment and spatial causal transition. Here, we focus on two challenges in spatio-temporal representation learning: (1) Eliminating temporal confounding bias. Existing models fail to handle causal relationships and certainly not eliminate the influence of confounders in temporal domain. Hence, the second challenge is to remove confounding bias and extract unbiased temporal representations. (2) Modelling spatial causal relationships. Limited by predefined graph structures, existing models are susceptible to non-causal spatial spurious correlations so it is significant to recover underlying spatial causal structure under causal constraints. First, we eliminate temporal spurious correlation based on backdoor adjustment. Then, we construct causal transition network for eliminating spatial spurious correlation. Finally, the downstream feature decoder applies the causal representation to downstream tasks. We tackle spatio-temporal representation learning tasks from a causal perspective and analyze the causes of spatial and temporal spurious correlation in observation data. To the best of our knowledge, this is the first attempt to apply causal theory to spatio-temporal representation learning. Experiments on two real-world datasets show that the proposed spatiotemporal representation learning method effectively avoids the interference of spurious correlations, enhances the stability of the model, and reduces the prediction error of nodes with sparse data in two downstream prediction tasks by 3% and 10% respectively.

Keywords spatio-temporal representation learning; structural causal model; spurious correlations; backdoor adjustment; causality relation

1 引 言

时空表示学习目的是将非结构化的高维时空数据转化为结构化的低维特征表示,使其能有效表征时空数据中潜在的规律与知识,并提高下游任务性能. 时空表示学习是时空数据挖掘领域的研究热点,其广泛应用于智能交通、城市规划、环境保护等领域.

时空数据的生成过程受到稳定因果结构的约束,因果结构不随数据分布的变化而改变,通过还原系统内各变量间的因果传递过程,可以有效提高从观测数据中提取的特征表示的鲁棒性. 时空相关性是由变量间的时空因果关系产生的,然而,当数据分布发生变化时,时空相关性会随之改变,主流时空表示学习方法通常将时空数据中的空间属性建模为图结构,时序属性建模为图上节点的时间序列,通过循环神经网络^[1-4]与图卷积神经网络^[5-9]建模观测数据中的时空相关性,提取特征表示并用于异常检测、时空序列预测等下游任务. 在深度神经网络的加持

下,时空表示学习方法的研究取得了巨大进展,然而由于数据集的不均衡性和模型的局限性,模型可能会学习到观测数据中虚假的时空相关性. 一方面,时域上的混杂因素导致历史数据与预测结果间存在非因果的虚假关联,使模型在建模过程受到时域虚假相关性的影响;另一方面,图卷积操作导致非因果关联节点间产生相关性,使模型在特征提取过程中受到节点间空域虚假相关性的影响,并最终导致生成了有偏的时空表示,使下游任务性能大幅下降.

在时域上,混杂因素使历史数据与预测结果间存在时域虚假相关性,而大部分模型在统一的参数空间下建模不同节点的时序相关性^[3-4],忽略了节点间的差异性,导致提取了有偏的时域表示. 图 1(a)展示了时空数据在时域上的因果图,箭头表示因果图中节点间的单向因果关系,其中 X 为历史数据, Y 为预测结果, C 为混杂因素, FT 为时域特征. 由于时空数据在时域上的自相关性,存在 $X \rightarrow Y$ 的因果路径. 时域特征 FT 是历史数据 X 在混杂属性 C 下的投影,并影响预测结果 Y , 因此存在 $\{X, C\} \rightarrow FT \rightarrow Y$ 的因果路径. 混杂因素 C 作为 X 与 Y 的共

因,通过 $X \leftarrow C \rightarrow FT \rightarrow Y$ 的路径使 X 与 Y 间产生虚假相关性. 数据集内混杂因素的取值分布具有不均衡性,使模型倾向于学习在数据集中具有主导地位的时序依赖,进而忽略在数据集中分布稀疏的时间序列所蕴含的依赖关系,并提取有偏的时域表示.

以出租车需求数据为例,图1(b)展示了两个典型区域的出租车需求变化过程. 对于工作区域,出租车需求呈现早、晚两个出行高峰的需求分布,而对于医院周边区域,出租车需求则呈现早、午、晚三个出行高峰的需求分布. 如果工作区域在数据集中比例较大,模型在训练过程中倾向于学习早高峰 \rightarrow 晚高峰的时序依赖,而忽略早高峰 \rightarrow 午高峰 \rightarrow 晚高峰的时序依赖,导致具有医院属性的区域在午高峰时段预测效果较差. 本例中,区域属性是出租车历史需求与未来需求间的混杂因素,而模型受其影响提取有偏表示的过程称为混杂偏倚(confounding bias).

在空域上,现存方法利用多层图卷积网络并基于无向图^[5,7]或双向扩散图^[6,8]聚合邻域节点特征并

构造空域表示,使非因果关联节点间产生空域虚假相关性. 一方面,无向图仅能建模对称的相关关系而非因果关系,另一方面,在多层图卷积操作后,双向扩散图中无因果关联的节点间可能建立虚假的空间相关性. 如图2(a)所示,在道路拥堵预测任务中,将路段看作节点,路段间的拥堵传递过程看作连边,其中深色表示路段拥堵,浅色表示路段畅通. 如图2(b)所示,在因果视角下,路段A与路段B及路段C存在因果关联. 基于消息传递机制的图卷积模型,会使路段B与路段C间产生相关性. 实际上路段B的拥堵不会导致路段C的拥堵,因此B与C间的相关性是虚假相关性. 在基于相关性的传统时空表示学习模型中,假如路段B由于路段D产生拥堵状态,根据B与C间的相关性,模型会误认为路段C也可能处于拥堵状态,进而给出错误的状态估计. 此外,节点间因果关系复杂多变,其中潜在的因果关联难以从先验知识(节点间交互量)中推出. 因此,如何从时空数据中挖掘节点间的潜在因果关联也是构造准确空域表示的关键.

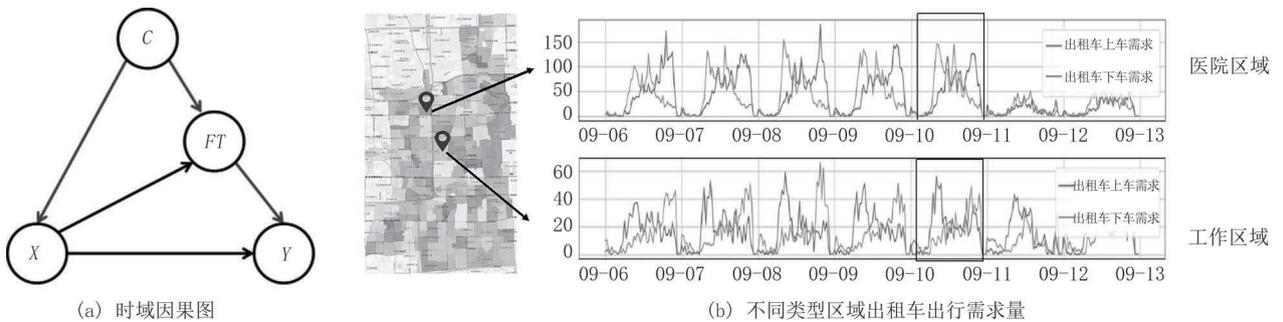


图1 时域虚假相关性图示

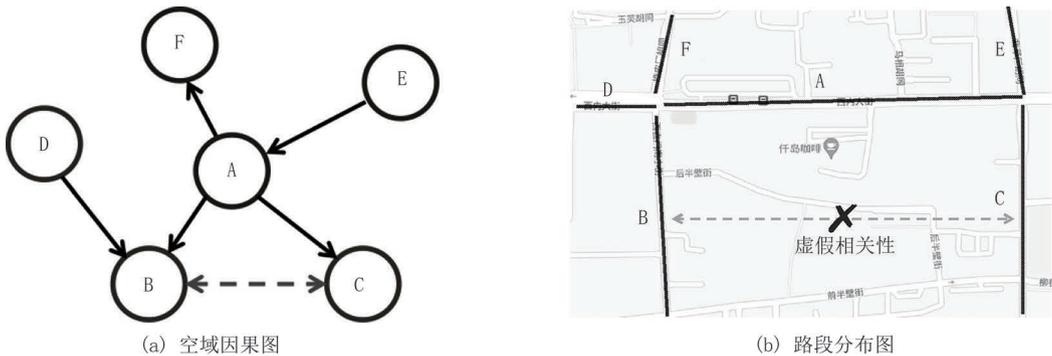


图2 空域虚假相关性图示

尽管现有时空表示学习方法在一定程度上建模了时空数据的相关性,但仍存在以下几点问题: (1)由于混杂因素的存在,当数据集分布不均匀的情况下,传统时空表示学习方法倾向于建模频繁出现

的时域虚假相关性,进而提取有偏的时域表示,造成混杂偏倚. (2)传统时空数据表示学习方法在不考虑节点间因果关联的相关性视角下提取空域特征,导致提取结果易包含因空域虚假相关性引入的非因

果关联,产生噪声。(3)现存方法难以从观测数据中还原各节点间的潜在因果关联,无法全面建模节点间的空域因果关系。总体而言,现有时空表示学习方法缺乏对时空数据生成过程中因果关系的建模,在虚假相关性的影响下,构建了有偏的时空表示,并降低下游任务性能。

为了解决传统上述问题,本文从因果视角下分析时空表示学习任务,提出了基于时域偏倚校正与空域因果传递的时空因果表示学习方法,包含以下三个主要部分:(1)本文提出基于时域偏倚校正的时域因果表示编码器,建模时空数据中的时序因果关系,基于后门调整(Backdoor Adjustment)在独立参数空间下校正时域因果表示,避免模型受到时域虚假相关性的干扰,提取无偏的时域因果表示;(2)本文提出基于空域因果传递的空域因果表示编码器,构建因果传递矩阵表征节点间的静态与动态因果关系,并设计因果传递网络建模节点间的因果信息传递过程,消除空域虚假相关性的影响并提取空域因果表示;(3)本文提出基于节点嵌入(Node Embedding)的潜在因果传递矩阵生成器,通过矩阵分解策略从输入数据中自适应学习动态节点嵌入,构建潜在因果关联矩阵,建模节点间潜在的动态因果依赖,从而全面建模节点间的空域因果关系。

本文的主要贡献分为以下三点:

(1)本文从因果视角下建模时空数据的生成过程,从数据集的不均衡性与模型的局限性两个角度分析观测数据中虚假相关性的成因,讨论了时域上的混杂因素及空域上的非因果关联对特征表示生成过程的影响;

(2)将因果图作为先验知识与深度学习模型融合,提出了基于时域偏倚校正与空域因果传递的时空因果表示学习方法 STCTN (Spatial-Temporal Causal Transition Network)。通过对时域特征进行去偏操作并建模空域上的因果传递过程,有效避免了虚假相关性对特征表示生成过程的影响,实现了无偏、稳定的因果表示;

(3)基于两个真实世界数据集进行了大量实验,与主流时空表示学习方法比较,本文模型有效提升了下游任务性能。实验结果表明了从因果视角下建模时空数据生成过程的正确性,验证了本文模型提取时空因果表示的有效性。

本文第1节介绍时空数据表示学习方法和因果表示学习方法的研究现状。第2节介绍本文所需的基础知识,即后门调整的基本概念和问题的形式化

定义。第3节介绍本文提出的时空因果表示学习方法。第4节通过实验验证了所提模型的有效性。最后总结全文。

2 相关工作

2.1 时空表示学习

传统的时空表示学习方法关注时空数据内蕴含的时空相关性。在空域上,为了表达节点间存在的非欧几里得关联,研究者们提出构建预定义图并基于图神经网络聚合邻域节点特征。Yu等人^[5]首次将GCN用于交通流预测领域,使用路段距离构建路网图,结合一维卷积和图卷积预测路段速度。Li等人^[6]关注到路段间具有上下游关系,提出双向扩散卷积用于有向图上的图卷积过程。Geng等人^[7]考虑到单个图在表征空间关系时的局限性,提出了多图卷积神经网络用于共享单车需求预测。Lv等人^[10]提出了一种基于网络嵌入的卷积结构以将空间信息融合在时序建模过程中,Song等人^[11]提出了一种基于图注意力网络的方法,并构建IST图以反映不同路段速度值的相关性。Qin等人^[12]提出了一种基于元学习的方法,在小样本学习场景下进行时空销量预测。Wang等人^[13]考虑到节点流量与节点间交互量的相关性,提出了一种多任务学习框架,在节点流量预测任务与节点交互量预测任务中共享信息以提高预测效果。考虑到节点间的动态空间依赖,研究者提出了一系列基于动态图的时空序列表示学习方法^[14-16],在节点属性随时间改变的同时,图结构也在不断发生变化。此类方法在图卷积过程中使非因果关联节点间产生虚假相关性,提取了有偏的空域表示。在时域上,大多数模型在统一的参数空间下基于卷积神经网络^[5,8]与循环神经网络^[6-7]建模时间相关性,忽略了节点间的差异性,Bai等人^[17]提出的自适应图卷积递归网络虽然为所有节点分配了独立参数空间,但是没有从因果视角下校正由混杂因素引发的混杂偏倚,导致提取有偏的时域表示。现有方法仅考虑了时空数据在统计上的相关关系,缺乏对因果关系的关注,本文则在因果视角下,避免虚假相关性的影响,并且通过一种端到端的训练方式学习节点间潜在的因果关联,从而生成稳定、有效的时空表示。

2.2 因果表示学习

因果表示学习是指通过建模数据中的因果关系,避免虚假相关性的干扰,将数据转化为符合物理规律的结构化表征^[18]。因果表示学习有效提高了模

型的泛化性及可解释性而被广泛应用于医疗、金融和社会学领域^[19-21]。近年来,许多研究者将因果表示学习用于计算机领域的相关问题。Lin等人^[22]在因果推断的视角下进行无监督视频异常检测任务,基于因果推理框架去除噪声伪标签对异常检测结果的影响。Liu等人^[23]通过解耦行人运动模式中物理法则、风格混杂和非因果虚假特征,保证行人运动表征的鲁棒性和可重用性,解决运动轨迹预测问题。Yang等人^[24]基于变分自编码器解耦图片中独立的语义因素,并通过干预操作生成符合因果关系的新图片。Yue等人^[25]通过后门调整去除小样本学习问题中预训练知识导致的样本特征与标签预测间的伪相关性,有效解决了在先验分布改变较大时,模型出现的“负迁移”问题。与上述研究针对的图片、自然语言等非时空数据相比,本研究关注时空数据中因果关系的建模,通过去除时域与空域上的虚假相关性,进而提取无偏的因果表示。

3 基础知识

本研究提出了一种因果视角下的时空表示学习方法,后门调整(Backdoor Adjustment)是结构因果模型理论中去除偏倚的重要手段^[26]。因此,本节介绍后门调整的相关概念并对时空表示学习问题进行形式化定义。

3.1 后门调整

如图3所示,给定包含原因变量 X ,结果变量 Y ,混杂变量 Z 的因果图 G ,因果学习的目的是求解 X 与 Y 之间的真实因果效应,即原因变量 X 的变化和取值如何影响结果变量 Y 。由于混杂变量 Z 是 X 与 Y 的共因,在两者间引入相关性,而在不同数据集中,由于 Z 的分布不同, X 与 Y 的相关性也不断变

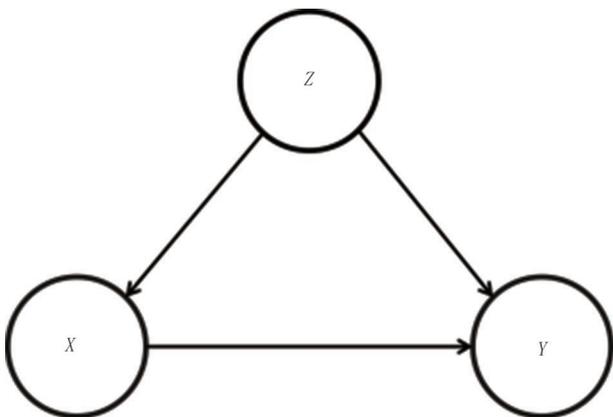


图3 因果图 G

化,称为虚假相关性。为了消除此类虚假相关性,需要切断 $X \leftarrow C \rightarrow Y$ 的非因果路径,在结构因果模型理论中,通常使用后门调整方法使混杂变量 Z 与原因变量 X 独立,以避免虚假相关性的干扰,其相关定义如下:

定义1. 后门路径. 给定因果图 G 中一对有序变量 (X, Y) , X 其中称为原因变量, Y 称为结果变量,如果存在一条路径,连接变量 X 与 Y ,同时具有指向原因变量 X 和结果变量 Y 的箭头,则称这条路径为 X 与 Y 之间的后门路径(Backdoor Path),后门路径在 X 与 Y 间建立了非因果的伪相关性。

定义2. 后门准则. 如果变量集合 Z 满足:(1) Z 中没有 X 的后代节点,(2) Z 阻断了 X 与 Y 之间每条后门路径,则称 Z 满足关于 (X, Y) 的后门准则(Backdoor Criterion)。

定义3. 后门调整. 如果变量集合 Z 满足后门准则,在 Z 可观测的情况下,变量 X 与 Y 间的因果效应可由后门调整公式求解:

$$P(Y|do(X)) = \sum_z P(Y|X=x, Z=z)P(Z=z) \quad (1)$$

其中, $do(X)$ 代表将 X 在不受其余因素影响下设置为某个值,称为干预(intervention)^[26],其通过切断了 X 与 Z 的因果路径使两者独立,进而消除了由后门路径建立的虚假相关性。

3.2 时空表示学习问题形式化定义

给定时空序列数据 $X \in \mathbb{R}^{N \times D \times S}$,其中 N 为节点数量, D 为特征维度, S 为时间步总数。 $X_t \in \mathbb{R}^{N \times D}$ 表示第 t 个时间步内所有节点的数据矩阵, $X^i \in \mathbb{R}^{D \times S}$ 表示节点 n_i 的时空数据矩阵。对于 Q 个时间步的历史观测数据 $X_{t+1:t+Q} \in \mathbb{R}^{N \times D \times Q}$,时空表示学习的目标是学习一个将时空数据映射到特征空间中的时空编码器 $f(\cdot): \mathbb{R}^{N \times D \times Q} \rightarrow \mathbb{R}^{N \times H}$, H 代表所提取的时空表示的特征维度。给定一个损失函数 $l(\cdot)$,对于所有时间步 t ,时空表示学习问题可形式化定义为

$$f^*, g^* = \operatorname{argmin}_{f, g} \sum_t l(g(f(X_{t+1:t+Q})), Y) \quad (2)$$

其中 f^*, g^* 为带有可学习参数的编码器和解码器, Y 则为下游任务的输出。例如,在序列预测任务中, $Y = X_{t+Q+1} \in \mathbb{R}^{N \times D \times 1}$,表示待预测的未来数据切片。

4 时空因果传递网络模型与建模流程

本文提出一种基于时域偏倚校正与空域因果传递的时空因果表示学习方法,即时空因果传递网络

STCTN (Spatial-Temporal Causal Transition Network), 整体网络模型主要包含三个主要部分: 时域因果表示编码器、空域因果表示编码器, 和下游任务解码器. 其主体结构如图4所示, 通过堆叠时域因果表示编码器与空域因果表示编码器, 逐时段提取时空因果表示, 并将最后一个时间步输出的因果表示矩阵作为输入数据的时空因果表示, 最后基于下游任务类型, 解码时空因果表示并进行端到端训练. 一方面, 现存模型大多利用循环神经网络及其变体建模时序相关性^[5-6, 10, 16-17], 在共享的参数空间下建模不同节点的时序演化模式. 本文在时序建模上沿用了循环神经网络变体GRU, 在此基础上利用因果学习方法消除混杂因素建立的虚假相关性. 首先, 对节点依据混杂因素取值分层, 并在独立参数空

间下构建各类型节点的特有预测模型, 最后基于后门调整提取无偏的时域因果特征; 另一方面, 现有方法利用多层图卷积网络并基于无向图^[5, 7]或双向扩散图^[6, 8]建模节点间的空间相关性, 其邻接矩阵往往是对称的, 仅能表达节点间的相关关系, 而多层图卷积网络在节点特征聚合过程中进一步引入了空间上的虚假相关性. 本文基于节点间交互强度构建非对称的显式因果传递矩阵, 并利用因果传递网络在仅执行一次节点间信息传递操作的前提下提取空域因果特征. 此外, 为了全面建模节点间的空域因果关系, 从观测数据中挖掘节点间潜在的因果关联, 本文基于矩阵分解方法学习节点的原因嵌入矩阵和结果嵌入矩阵, 在无需先验知识的情况下自适应生成潜在因果传递矩阵.

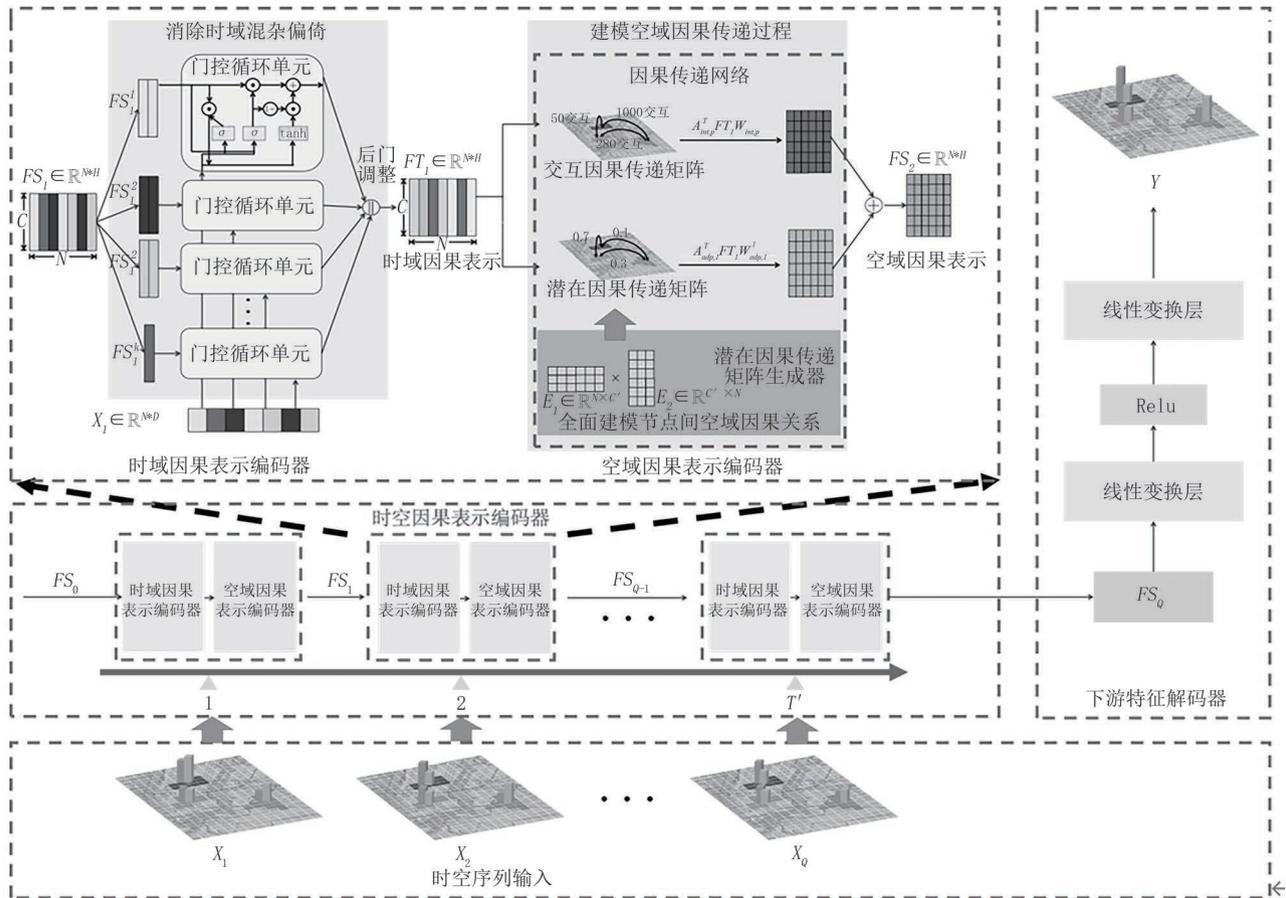


图4 STCTN模型整体结构图

4.1 基于时域偏倚校正的时域因果特征编码

传统时空表示学习方法^[5-7, 10-13]只关注历史数据与未来数据间的统计相关性, 忽略了对因果关系的建模, 在混杂因素的影响下, 模型倾向于提取一般性的时域特征, 提取有偏表示, 这种现象称为混杂偏倚. 为了校正偏倚并消除其带来的影响, 本文基于

后门调整方法提取无偏的时域因果特征.

现存预测模型基于历史数据 X 与预测结果 Y 的相关性, 采用概率模型 $P(Y|X)$ 来建模时空序列数据中的时序依赖, 由于混杂因素 C 的存在, X 与 Y 之间存在 $X \leftarrow C \rightarrow FT \rightarrow Y$ 的后门路径, 导致所提取的时域特征受到其引入的虚假相关性的影响, 若将

混杂因素 C 按照取值分层为 $C = \{c^1, c^2, \dots, c^p\}$, $P(Y|X)$ 可被分解为

$$P(Y|X) = \sum_k^p P(Y|X, FT = \mathcal{F}(X, c^k)) P(c^k|X) \quad (3)$$

其中 $\mathcal{F}(\cdot, \cdot)$ 为时域表示的提取模型, $P(c^k|X)$ 表示历史数据样本 X 和各类混杂 c^k 共同出现的概率, 在大多数情况下, 由于数据集的不均衡性, 历史数据样本 X 与各混杂因素 C 不能以均等概率结合, 导致模型所提取的特征 FT 倾向于表达频繁出现的一般性时序依赖, 即存在混杂偏倚. 为了消除历史数据 X 与预测结果 Y 间的虚假相关性, 本文基于后门调整公式, 使用 $P(Y|do(X))$ 替换 $P(Y|X)$ 以建模时序因果关系. 如图5所示, 此操作会切断 X 与 C 间的因果路径, 令 X 与 C 独立, 此时 $P(c^k|X) = P(c^k)$, 进而切断 $X \leftarrow C \rightarrow FT \rightarrow Y$ 的后门路径, 消除混杂因素 C 带来的虚假相关性, $P(Y|do(X))$ 可由下式分解:

$$P(Y|do(X)) = \sum_k^p P(Y|X, FT^k = \mathcal{F}_k(X, c^k)) P(c^k) \quad (4)$$

为了使历史数据样本与混杂因素公平结合, 假设混杂因素 C 服从离散均匀分布, 即 $C \sim U(p)$, p 为混杂因素的可取值总数.

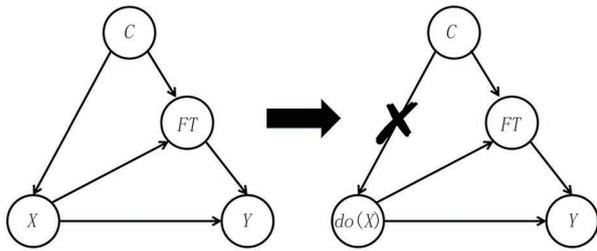


图5 后门调整

本文使用门控循环单元(GRU)作为时序因果特征提取模型 $\mathcal{F}_k(X, c^k)$, 时间步 t 内混杂因素取值为 c^k 的节点 i 的时域因果特征 FT_t^i 计算过程如下式:

$$\begin{aligned} z_t^i &= \sigma(W_{(z)}^k X_t^i + U_{(z)}^k FS_{t-1}^i) \\ r_t^i &= \sigma(W_{(r)}^k X_t^i + U_{(r)}^k FS_{t-1}^i) \\ \tilde{H}_t^i &= \tanh(W^k X_t^i + (r_t^i \odot U^k) FS_{t-1}^i) \\ FT_t^i &= z_t^i \odot \tilde{H}_t^i + (1 - z_t^i) FS_{t-1}^i \end{aligned} \quad (5)$$

其中, W, U 为可学习权重, \odot 代表矩阵的逐元素乘法, $\sigma(\cdot)$ 代表 sigmoid 激活函数, FS_{t-1}^i 代表上一时间步内提取的空域因果特征, FS_0^i 初始化为零矩阵.

最后, 合并各节点提取的时域因果特征, 得到时域因果特征矩阵 FT_t 如下式:

$$FT_t = \text{BatchNorm}\left(\left\| \frac{1}{p} FT_t^{1} \right\| \left\| \frac{1}{p} FT_t^{2} \right\| \dots \left\| \frac{1}{p} FT_t^{p} \right\| \right) \quad (6)$$

其中 $\|$ 代表在节点维度上的矩阵拼接操作, $\text{BatchNorm}(\cdot)$ 为批归一化操作, 防止梯度消失, 加快模型收敛.

本文所提方法与现存方法^[5-7,10-13] 有两点不同: 首先, 为不同类型节点的时空序列分配独立的模型 \mathcal{F}_k , 提取差异化的时序因果特征; 其次, 为了使模型不受混杂偏倚的影响, 利用后门调整方法切断了 C 与 X 间的因果路径, 使历史数据样本与混杂因素等概率结合, 最终提取了无偏的时序因果特征.

4.2 基于空域因果传递的空域因果特征编码

现存模型基于图卷积网络建模节点空间相关性^[5-7,14-17], 在非因果相关节点间引入虚假关联, 为了提取无偏的空域因果特征, 本文在构建因果传递矩阵的基础上, 提出因果传递网络(Causal Transition Network)建模节点间的空域因果传递过程. 因果传递网络包含三部分: 首先, 基于先验知识构建显式因果传递矩阵, 表征节点间的显式空域因果传递关系; 其次, 基于节点嵌入构建潜在因果传递矩阵, 表征节点间的隐式空域因果传递关系; 最后, 依据因果传递矩阵计算传递后因果特征.

图6所示的因果传递过程所对应的因果传递矩阵 A 可表示为

$$A = \begin{pmatrix} 0 & 0 & 0.2 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0.2 \\ 0.3 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0.1 & 0 & 0 \\ 0.5 & 0.2 & 0 & 0 & 0 \end{pmatrix} \quad (7)$$

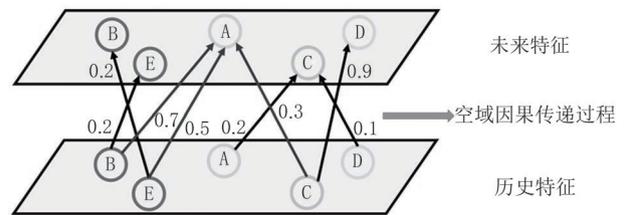


图6 空域因果传递过程

矩阵第 i 行第 j 列的元素表示以节点 j 历史特征为原因变量, 节点 i 未来特征为结果变量的因果关联强度. 对于节点 A , 受到节点 B, C, E 的因果效应的影响, 其经由因果传递过程更新后特征 x_A' 可由下式计算:

$$x_A' = 0.7 \times x_B + 0.3 \times x_C + 0.5 \times x_E \quad (8)$$

其中 x_B, x_C, x_E 为节点 B, C, E 的更新前特征, 假设更

新后全部节点特征矩阵为 x' , 因果传递网络可表示为

$$x' = A^T x W' \quad (9)$$

其中 A^T 表示因果传递矩阵的转置, W' 为用于调整特征维度的可学习参数矩阵.

为了建模节点间的因果关系, 构建了两类因果传递矩阵, 定义如下:

(1) 交互因果传递矩阵: 交互行为在节点间产生因果关联, 而在不同时间段内, 节点间的交互因果强度不同, 例如, 在出租车需求预测任务中, 出发地的需求量改变会同步引起到达地需求量的改变. 在早高峰时段, 居住区域的高出发需求可能导致工作区域的高到达需求, 而在晚高峰时段, 工作区域的高出发需求可能导致居住区域的高到达需求. 因此本文构建交互因果传递矩阵, 设时段内节点 i 到节点 j 的交互量为 $\lambda_{m,i,j}^{i,j}$, 构建时段 m 内交互因果传递矩阵如下式:

$$\tilde{A}_{int,m} = \begin{cases} \lambda_{m,i,j}^{i,j} & i=j \\ 0 & i \neq j \end{cases}$$

$$A_{int,m} = D^{-1} \tilde{A}_{int,m} \quad (10)$$

其中 D 为矩阵 $\tilde{A}_{int,m}$ 的度矩阵, 若时间步 t 属于时段 m , 则空域因果传递过程可表示为

$$FS_{int,t} = A_{int,m}^T FT_t W_{int,m} \quad (11)$$

其中 FT_t 为时间步 t 下提取的时域因果特征, $A_{int,m}^T$ 为时间段 m 内交互因果传递矩阵的转置, $W_{int,m}$ 为调整特征通道的可学习参数矩阵, $FS_{int,t}$ 为时间步 t 下的由交互因果传递矩阵提取的空域因果特征.

(2) 潜在空间因果传递矩阵: 基于先验知识构建的预定义矩阵难以表达节点间的潜在因果关联, 为了使模型从数据中自主构建节点间的潜在因果传递机制, 本文在反向传播过程中端到端(End-To-End)地生成因果传递矩阵 A_{adp} , 考虑到节点间因果关联的动态性, 在不同时间步内, 学习独立的矩阵 $A_{adp,t}$. 在时间步 t 内, 其空域因果传递过程可表示为

$$FS_{adp,t} = A_{adp,t}^T FT_t W_{adp,t} \quad (12)$$

其中 FT_t 为时间步 t 下提取的时域因果特征, $A_{adp,t}^T$ 为时间步 t 下交互因果传递矩阵的转置, $W_{adp,t}$ 为调整特征通道的可学习参数矩阵, $FS_{adp,t}$ 为时间步 t 下的由自适应因果传递矩阵提取的空域因果特征.

在节点数量较多的情况下, 直接通过梯度下降拟合潜在因果传递矩阵会导致模型的过参数化和较大的计算负载, 因此本文将潜在因果传递矩阵分解, 如下式:

$$A_{adp,t} = \text{softmax}(E_{1,t} E_{2,t}^T) \quad (13)$$

其中, $E_{1,t} \in \mathbb{R}^{N \times C_t}$ 称为时间步 t 下节点的原因嵌入矩阵, $E_{2,t} \in \mathbb{R}^{N \times C_t}$ 称为时间步 t 下节点的结果嵌入矩阵, 则空域因果传递过程可改写为

$$FS_{adp,t} = \left(\text{softmax}(E_{1,t} E_{2,t}^T)^T \right) FT_t W_{adp,t} \quad (14)$$

通过上述过程, 可学习参数量由 $N \times N \times T$ 下降到 $2 \times N \times C_t \times T$ ($N \gg C_t$), 同时有效减低了计算负载, 并通过降低参数量防止模型过拟合.

最后, 融合两类因果传递过程中提取的空域因果特征, 如下式:

$$FS_t = \text{AGG}(FS_{int,t}, FS_{adp,t}) \quad (15)$$

其中 $\text{AGG}(\cdot)$ 为聚合函数, 本文使用求和函数作为聚合函数, 即将特征矩阵对应位置求和以生成融合特征. 所提取的空域因果特征作为隐层特征参与下一时间步中时域因果表示的生成.

与现存预测模型^[5-7, 14-17]不同, 本文通过构建节点间的因果传递矩阵, 在因果视角下建模节点间的空域因果传递过程, 有效避免了现存模型中由图卷积网络引入的节点间非因果关联. 与 ST-MGCN 模型^[7]中使用同一无向交互图不同, 本文利用各时间段下节点间的差异性交互模式构建了时间特定性交互因果传递矩阵, 在各时段下构建非对称因果传递矩阵, 进而表征节点间交互因果关联. 此外, 本文提出一种潜在因果关联矩阵生成方法, 从数据中挖掘节点间潜在因果关联, 并通过矩阵分解减少参数量, 降低计算复杂度.

4.3 下游任务解码

下游任务解码器旨在将最后一个时间步 Q 下的融合特征 FS_Q 映射到任务所需的特征空间上, 并通过端到端的方式训练编码器并提取时空因果表示, 本文使用堆叠的全连接层作为解码器. 使用 ReLU 作为激活函数. 解码结果 Y 可表示为

$$Y = W_2(\text{ReLU}(W_1 FS_Q) + b_1) + b_2 \quad (16)$$

其中 W_1, b_1, W_2, b_2 为全连接层的权重和偏置. 解码器的输出与下游任务类型相关, 在预测任务中, $Y = X_{Q+1} \in \mathbb{R}^{N \times D \times 1}$, 表示待预测的未来数据切片.

5 实验与分析

5.1 实验数据及参数设置

本文在 NYC-Taxi, XC-Taxi 两个城市出租车订单数据集与上进行实验, 将提取的时空因果表示用于时空序列预测任务. 表 1 给出了原始数据集所对应的详细数据信息.

表1 实验数据集

数据集	时间跨度	订单数量	区域数量	POI种类	POI数量
NYC-Taxi	2016/04/01-2016/06/30	30 735 780	90	学校、医院、超市、公司、餐馆、银行	4867
XC-Taxi	2021/05/01-2021/12/31	89 806 692	263	学校、医院、超市、公园、餐馆、银行	2596

NYC-Taxi数据集中包含美国纽约曼哈顿区自2016年4月1日至2016年6月30日的出租车订单数据,研究区域被划分为90个子区域.XC-Taxi数据集中包含北京市西城区自2021年5月1日至2021年12月31日的出租车订单数据,研究区域被划分为263个子区域.出租车订单数据中包含上车时间、下车时间、上车位置、下车位置等信息,将全部订单按照30分钟间隔转化为各区域出租车上车需求与下车需求,删除了其中的异常订单(出行距离过长和出行时间过长的订单).NYC-Taxi的需求量数据共包含4362个样本,输入的时间序列长度为 $Q=6$,输入维度为 $6 \times 90 \times 2$,输出维度为 $1 \times 90 \times 2$.XC-Taxi的需求量数据共包含11754个样本,输入的时间序列长度为 $Q=6$,单个输入维度为 $6 \times 263 \times 2$,输出维度为 $1 \times 263 \times 2$.

本文使用前60%的样本作为训练集,其后20%作为验证集,剩余20%用作测试集.对于时域因果特征编码器模块和空域因果特征编码器的隐藏单元维度,本文在参数集合 $\{16, 32, 64\}$ 中调优,并最终设置隐藏单元维度 $H=32$.对于原因嵌入向量和结果嵌入向量维度,本文在参数集合 $\{5, 10, 15, 20\}$ 中调优,并最终设置嵌入向量维度 $C_i=10$.在出租车需求预测问题中,区域属性作为混杂因素导致了混杂偏倚,本文基于谱聚类算法将子区域划分为具有不同区域属性的聚簇,并以此为依据对混杂因素(区

域属性)分层.为了选择最合适的聚簇数,本研究使用轮廓系数(Silhouette Coefficient)作为聚类效果的评价标准,轮廓系数越大,聚类效果越好.如图7所示,当聚簇数 p 为7和4时,聚类效果最好.本文模型的激活函数为ReLU,通过Adam^[27]优化器对网络进行优化,初始学习率和权重衰减值分别设置为 $1e-3$ 和 $1e-4$,并使用SmoothL1作为损失函数.本文基于PyTorch搭建模型,并在NVIDIA 3090 GPU上迭代训练200个轮次,当模型在验证集上的表现连续50轮没有提高时终止训练,选择在验证集上表现最好的模型作为最终模型.

5.2 评价指标

本文采用时空序列预测领域常用的评价指标平均绝对误差(Mean Absolute Error, MAE),均方根误差(Rooted Mean Square Error, RMSE),和平均百分比误差(Mean Average Percentage, MAPE)作为评估模型性能的指标,计算式为

$$\begin{aligned} \text{MAE} &= \frac{1}{z} \sum_{i=1}^z |\hat{Y}_i - Y_i| \\ \text{RMSE} &= \sqrt{\frac{1}{z} \sum_{i=1}^z (\hat{Y}_i - Y_i)^2} \\ \text{MAPE} &= \frac{1}{z} \sum_{i=1}^z \frac{|\hat{Y}_i - Y_i|}{Y_i} \end{aligned} \quad (17)$$

其中, \hat{Y}_i 代表预测结果, Y_i 代表真实值, z 代表样本个数.

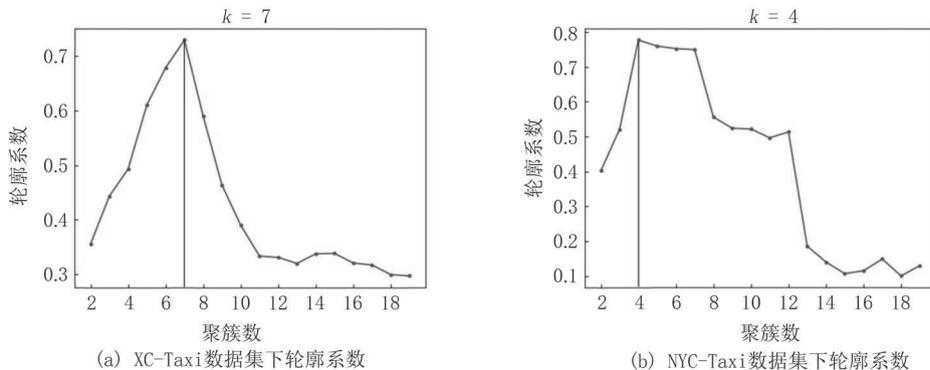


图7 不同数据集下的轮廓系数

5.3 基准模型与实验对比

将本文所提方法STCTN与主流时空序列预测

方法进行比较,其中时间序列预测方法包括LSTM和GRU,此类方法仅考虑时序相关性;对于同时考

虑时空维度的预测模型,本文依据其所使用的图的类型进行划分,其中基于静态图的方法包括STGCN^[5], HGCN^[28],基于自适应图的方法为GraphWavenet^[8]; 基于动态图的方法包括AGCRN^[17], DGCRN^[16], DMSTGCN^[29]. 为了保证实验的公平性,本文对以上方法均进行了参数调优,并使用与本文所提方法STCTN相同的训练集、验证集、测试集以及输入时间序列长度进行实验验证. 各模型参数设置如下:

LSTM:设置其隐藏单元维度为64.

GRU:设置其隐藏单元维度为64.

STGCN:使用区域间距离构建图结构,并设置隐藏单元维度为64.

HGCN:使用区域间距离构建图结构,并将其中图卷积神经网络和卷积神经网络的隐藏单元维度设置为32,节点嵌入维度设置为10.

GraphWavenet:节点嵌入维度设置为20,隐藏单元维度设置为32.

AGCRN:隐藏单元维度设置为64,切比雪夫多项式阶数设置为2,节点嵌入维度设置为10.

DGCRN:隐藏单元维度设置为96,节点嵌入维度设置为40.

DMSTGCN:该模型为不同时间段构造专属图结构,将一天分解为48个时间段,隐藏单元维度设置为32.

表2展示了对比实验结果. 本文在两个数据集上进行了五次重复实验,并计算各项指标的均值作为模型最终预测效果. 从表中可以看出,本文所提方法STCTN在两个数据集的各评价指标上均达到了最佳效果. 图8展示了各模型在不同数据集上的预测方差(由五次重复实验中MAE的方差表示),方差越小,模型稳定性越强.

表2 对比试验结果

模型	XC-Taxi			NYC-Taxi		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
LSTM	3.5498	6.1021	30.7581%	9.4459	16.5967	17.6489%
GRU	3.5586	6.1246	31.0366%	9.5328	16.7341	17.9292%
STGCN	3.0191	5.2800	26.5357%	8.5629	15.6340	15.5858%
HGCN	3.0770	5.3910	27.4320%	9.3608	17.4308	16.7008%
GraphWavenet	2.9805	5.1019	26.0979%	8.3034	14.8816	15.2806%
AGCRN	3.0289	5.3207	26.5788%	8.3237	15.0525	15.2555%
DGCRN	2.9618	5.1980	26.2017%	8.1999	14.6100	15.1004%
DMSTGCN	2.9838	5.1373	26.1987%	8.1302	14.5972	15.1566%
STCTN	2.9294	5.0986	26.0869%	8.0813	14.5303	14.9117%

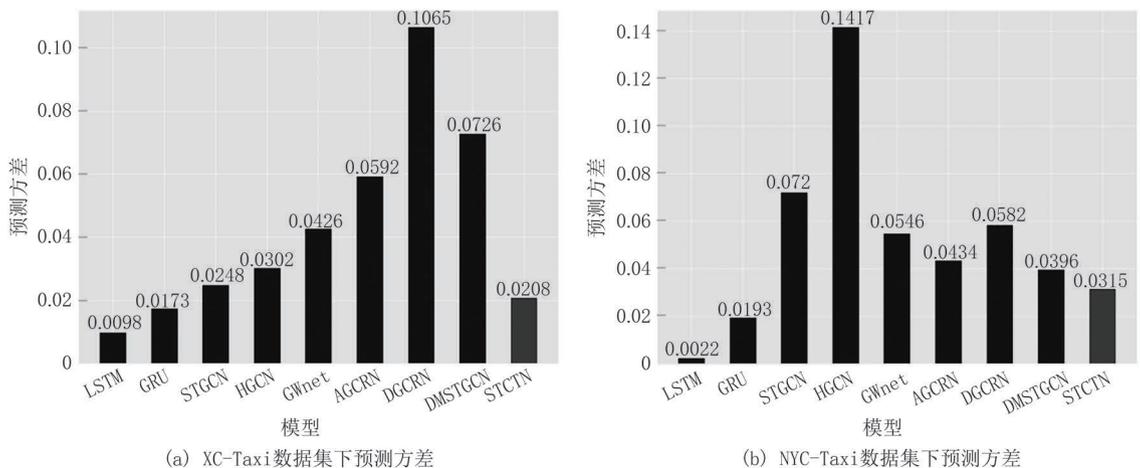


图8 各模型在XC-Taxi和NYC-Taxi数据集下的预测方差

其中,由于时间序列预测方法LSTM与GRU的模型较为简单,模型稳定性最高. 但是其仅考虑

了时序相关性,而忽略了节点间空间依赖的建模,在两个数据集上的预测误差最大.

在时域上, STGCN、HGNC、GraphWavenet、DMSTGCN基于卷积神经网络建模时序相关性, 而DGCRN基于循环神经网络建模时序相关性, 上述模型均在共享参数空间在提取时域特征, 在数据集不均衡的情况下, 倾向于提取有偏的时域表示. AGCRN为各节点分配了独立的参数空间, 但是没有从因果视角下校正各类节点的特征表示, 无法提取无偏的时域特征表示.

在空域上, 基于静态图的STGCN和HGNC没有考虑到节点间的多类因果关联, 虽然其表现要优于时间序列预测模型, 但是没有全面建模节点间的交互行为和潜在动态依赖, 预测效果较GraphWavenet、DGCRN、DMSTGCN等基于自适应图和动态图的预测模型仍有一定差距. 此外, STGCN与HGNC基于无向图卷积建模空间相关性, GraphWavenet与DGCRN基于双向扩散图卷积建模空间相关性, 基于消息传递机制的图卷积弱化了节点间固有的因果关系, 在空域特征提取过程中引入虚假相关性, 在一

定程度上降低了模型的精度与稳定性.

与上述方法相比, STCTN基于后门调整在独立参数空间下校正了时域因果特征提取过程中的混杂偏倚, 并通过构建因果传递矩阵建模了节点间的因果传递过程, 避免了非因果的相关性对模型提取的特征表示的影响, 模型稳定性更强, 预测表现更为优异.

5.4 不同类型节点实验结果

为了探讨对混杂因素取值不同的节点预测效果的差异, 本文在两个XC-Taxi和NYC-Taxi数据集上针对具有不同功能属性的区域进行实验, 表3与表4分别展示了两个数据集上各具有不同功能属性的区域数量以及主流预测模型在具有不同功能属性区域上的预测效果, 以各模型的平均绝对误差(MAE)作为评价模型表现的指标.

如表3与表4中所示, STCTN在各类区域中表现较对比模型均有所提升, 证明了模型所提取的时空因果表示的鲁棒性. 此外, 从表中可以发现, 数据集中存在各类型区域数量不均衡情况.

表3 XC-Taxi中各聚簇信息及模型预测表现(MAE)

模型	办公区域	购物区域	金融区域	医疗区域	餐饮区域	学校区域	其他区域
STGCN	2.9266	3.2931	3.5911	2.1071	3.0027	2.3704	2.3028
HGNC	2.9982	3.3515	3.6368	2.1346	3.0458	2.4586	2.3299
GraphWavenet	2.8998	3.2617	3.4985	2.0869	2.9845	2.3506	2.2978
AGCRN	2.9347	3.3154	3.6043	2.0948	3.0221	2.3645	2.2969
DGCRN	2.8841	3.2436	3.4809	2.0785	2.9508	2.3382	2.3021
DMSTGCN	2.9179	3.2317	3.4979	2.1020	2.9796	2.3834	2.3094
STCTN	2.8428	3.1844	3.4524	2.0327	2.9295	2.3073	2.2235
区域数量	85	28	59	21	47	18	5

表4 NYC-Taxi中各聚簇信息及模型预测表现(MAE)

模型	娱乐区域	医疗区域	办公区域	其他区域
STGCN	5.8466	9.8930	10.3553	1.8807
HGNC	6.2845	10.7588	11.2694	2.4356
GraphWavenet	5.7360	9.4957	9.9938	1.8295
AGCRN	5.7304	9.5817	10.0123	1.9067
DGCRN	5.6873	9.2054	9.9097	1.9658
DMSTGCN	5.7164	9.2776	9.7857	1.9069
STCTN	5.5526	8.9154	9.7074	1.6258
区域数量	32	9	47	2

例如, 在XC-Taxi数据集中, 学校区域分布较为密集, 而购物区域分布则较为稀疏, 此时区域属性作为时域上的混杂因素, 在模型预测过程中引入了混杂偏倚, 本文引入了后门调整避免这种由于数据不均衡所提取的有偏表示. 对于数据稀疏区域, 本文所提方法在XC-Taxi数据集与NYC-Taxi数据集

中误差分别降低了3%和10%. 图9展示了在各类区域簇中STCTN与基准模型的表现对比, 其中纵坐标为归一化后的平均绝对误差(MAE), 横坐标为按类内区域数由大到小排序的区域聚簇. 随着簇内区域数的减少, STCTN的预测效果较基线模型的提升更为明显. 主流的预测模型在共享的参数空间下建模时序依赖, 倾向于提取办公区等数据丰富区域的时序特征, 导致其难以表达购物区域、学校区域等数据稀疏区域的出租车出行模式, 在此类区域上预测效果不佳. 本文所提模型则在私有参数空间下分别提取不同类型区域的时序特征, 从因果视角下公平地关注各类型区域出租车需求的时域因果关联, 生成无偏的时域因果表示, 当某种类型区域数据较为稀疏时, 仍能取得较好的预测结果.

为了进一步说明本文模型对数据稀疏类型节点

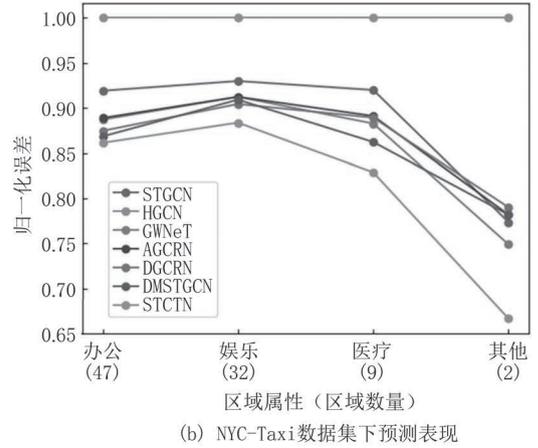
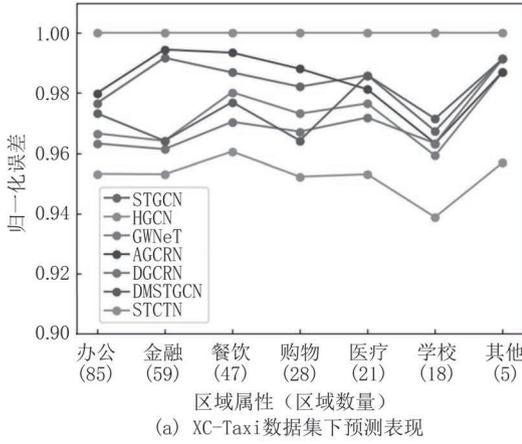


图9 各模型不同类型区域簇下的预测表现

的有效性,图10展示了各模型在北京市儿童医院周边区域的出租车到达需求预测结果.从图中可以看出,STCTN模型在早高峰时段(浅色区域)对医院周边区域的到达需求给出了准确的预测.对于医院

特有的午高峰时段(深色区域),基线模型的预测效果较差,而STCTN通过时序上的特征去偏操作,有效提取了数据稀疏类型区域的无偏时序因果特征,并给出了准确的预测结果.

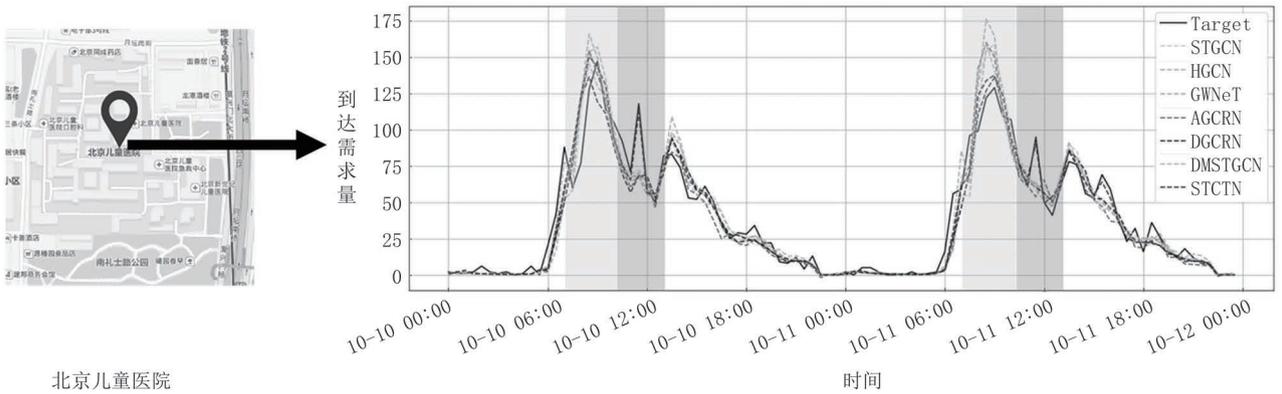


图10 数据稀疏类型区域(北京儿童医院)出租车需求预测结果

5.5 因果传递矩阵可视化

图11展示了交互因果传递矩阵的可视化结果,以图11(a)中的81号学校区域与93号居住区域为例,两者间存在时间特定性的交互因果关联,如图11(b)、图11(c)及图11(d)所示,在早高峰时段,因果传递方向为居住区域指向学校区域,而在晚高

峰时段,因果传递方向为学校区域指向居住区域,在周末时段,两者间则没有显著的因果关联,这说明区域间的因果关联在不同时段下是具有差异甚至完全相反的.基于图卷积的模型一般通过统计节点间的交互量生成在全部时段内统一的无向交互图,既无法建模区域间的有向因果关联,又不能表达在不同

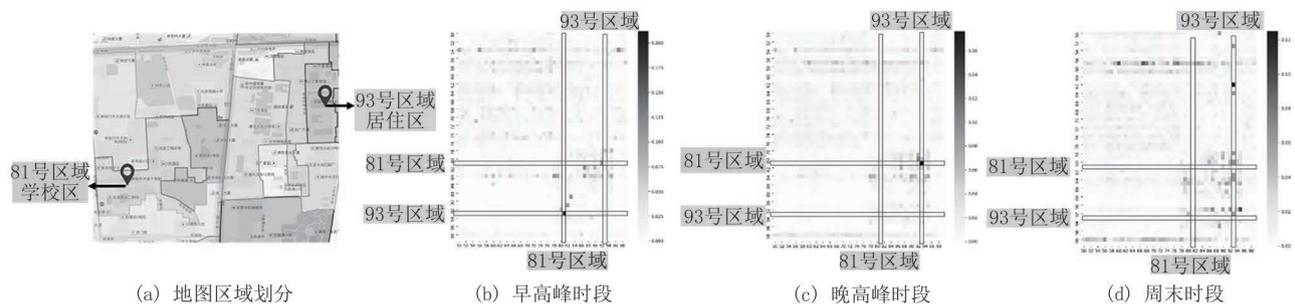


图11 交互因果传递矩阵

时段内节点间的差异性因果关联,导致其难以提取有效的空域特征表示.

图12展示了部分区域潜在因果传递矩阵的可视化结果,模型捕捉到了出租车需求在西城区主要医院间的因果传递关系.这种因果关系的产生是由于医院的分流效应,积水潭医院附近的出租车需求

量的上升意味着就诊人数的上升,一旦超出其承载力,就医人群可能会选择北医一院作为替代,进而引起周边区域出租车需求量的上升,而儿童医院与北医一院由于优势科室不同,可替代性较低,因此因果关联较弱,这进一步验证了模型通过端到端训练有效挖掘了节点间的潜在因果关联.

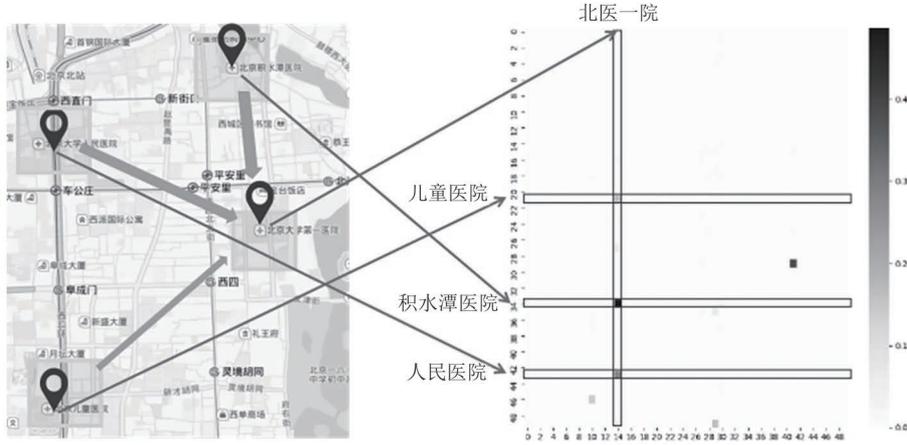


图12 潜在因果传递矩阵

5.6 组件有效性测试

为了验证STCTN中关键组件对预测结果的影响,本研究构建如下变体模型:

(1)base model,该模型使用单个GRU建模时序依赖,并使用无向距离图和图卷积神经网络建模空间依赖,代表仅捕捉时空相关性的模型.

(2)w/o inter,该模型删去了空域因果特征编码器中的时间特异性交互因果传递矩阵,以此探讨高频交互节点间因果传递对模型性能的影响;

(3)w/o adp,该模型删去了空域因果特征编码器中的潜在因果传递矩阵,以此探讨潜在关联节点间因果传递对模型性能的影响;

(4)static adp,该模型将空域因果特征编码器中各个时间步下使用同一潜在因果传递矩阵,以此探讨交互因果传递矩阵的时间特异性对模型性能的影响;

(5)single GRU,该模型令时域因果特征编码器中的所有节点共享同一个编码器,这种方式会导致模型提取有偏的时序特征,以此探讨在缺乏时域偏倚校正时模型的性能.

表5与图13展示了上述变体与本文所提STCTN在XC-Taxi上的实验结果对比.从表中可以看出,两类因果传递矩阵在不同程度上提高了预测的准确度,其中交互因果传递矩阵对模型性能的

表5 STCTN及变体模型预测结果

模型	MAE	RMSE	MAPE
base model	3.1917	5.6553	28.3067%
w/o inter	2.9976	5.2062	26.4998%
w/o adp	2.9781	5.1712	26.5586%
static adp	2.9538	5.1576	26.4070%
single GRU	2.9830	5.2092	26.7516%
STCTN	2.9294	5.0986	26.0869%

影响较大,这表明时间特定性的交互因果传递矩阵有效捕捉了由于差异化交互行为产生的因果关联.此外,STCTN的预测表现优于static adp变体,这说明潜在因果关联矩阵的动态性与时空数据中蕴含的动态因果关联一致,有效建模了节点间潜在的空域因果关联.

为了验证时域上混杂偏倚的特征表示生成过程的影响,仿照传统时空表示学习模型,在共享数空间下提取时域特征,设计了single GRU变体.与STCTN相比,single GRU预测表现大幅下降,说明时域混杂偏倚会导致模型生成有偏的时域表示.

相比于仅建模时空相关性的base model变体,本文所设计的时域因果特征编码器和空域因果特征编码器均能有效提高预测精度,证明了从因果视角下建模时空表示学习任务,并构建无偏因果表示的必要性.

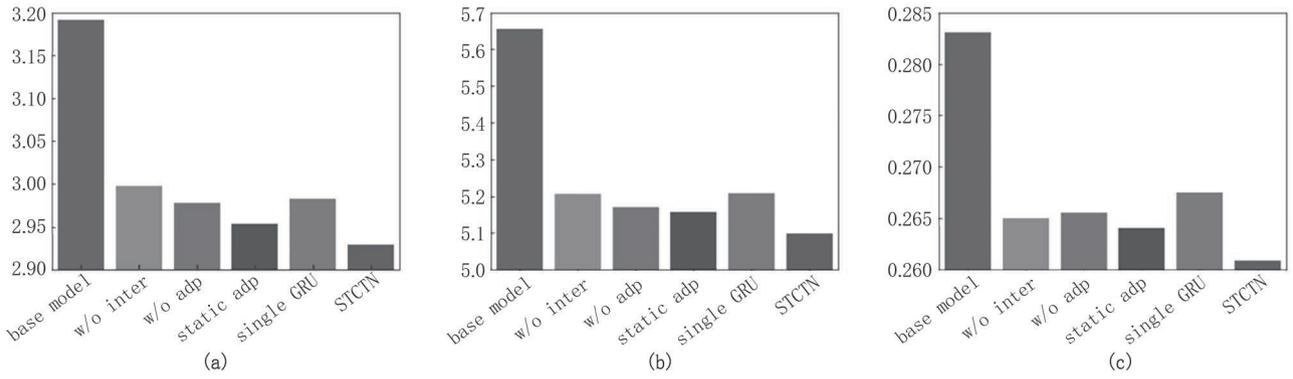


图13 STCTN及变体模型测试结果

为了验证空域上非因果关联对的特征表示生成过程的影响,本文仿照传统时空表示学习模型中的多层图卷积网络,增加空域因果表示器中的因果传递网络层的层数.如图14所示,随着因果传递网络层数的增加,模型的预测误差不断变大.这说明,多层图卷积操作会引入节点间的非因果关联,并导致模型生成有偏的空域表示.

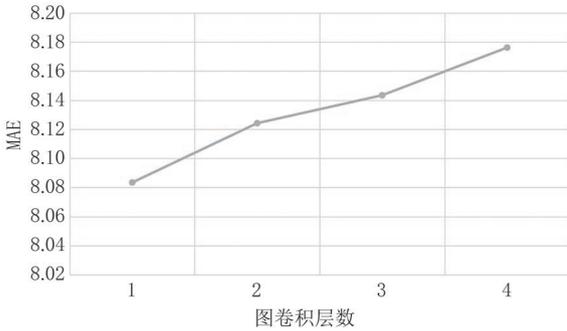


图14 因果传递网络层数对模型预测性能的影响

6 总 结

针对传统时空表示学习方法受到虚假相关性干扰而生成有偏表示的情况,本文提出了一种基于时域偏倚校正和空域因果传递的时空因果表示学习方法STCTN以建模时空序列内蕴含的因果关系.在时域上,通过混杂因素取值对时空序列分层,基于后门调整公式去除混杂因素引起的混杂偏倚,提取无偏的时域因果特征.在空域上,构建因果传递矩阵以表征节点的因果关联,并通过动态节点嵌入挖掘节点间的潜在因果关联,进一步提出因果传递网络

建模节点间的空域因果传递过程.通过在两个真实世界数据集上与多种时空表示学习方法的比较,表明了从因果视角下建模时空数据生成过程的正确性,验证了本文模型提取时空因果表示的有效性.本方法的局限性在于需要人为定义数据集的混杂因素,而这种定义方法不够全面,可能存在未被识别的混杂因素,导致模型不能完全消除虚假相关性的影响,未来我们将对此进行优化,以期自动发现观测数据中的混杂因素.

参 考 文 献

- [1] Lai Y, Zhang K, Lin J, et al. Taxi demand prediction with LSTM-based combination model//Proceedings of the IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom). Xiamen, China, 2019: 944-950.
- [2] Xu J, Rahmatizadeh R, Bölöni L, et al. Real-time prediction of taxi demand using recurrent neural networks. IEEE Transactions on Intelligent Transportation Systems (TITS). 2017, 19(8): 2572-2581.
- [3] Shu P, Sun Y, Zhao Y, et al. Spatial-temporal taxi demand prediction using LSTM-CNN//Proceedings of the 16th International Conference on Automation Science and Engineering (CASE). 2020: 1226-1230.
- [4] Kuang L, Yan X, Tan X, et al. Predicting taxi demand based on 3D convolutional neural network and multi-task learning. Remote Sensing. 2019, 11(11): 1265.
- [5] Yu B, Yin H, Zhu Z. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting//Proceedings of the 27th International Joint Conference on

- Artificial Intelligence (IJCAI). Stockholm, Sweden, 2018: 3634-3640.
- [6] Li Y, Yu R, Shahabi C, et al. Diffusion convolutional recurrent neural network; data-driven traffic forecasting// Proceedings of the 6th International Conference on Learning Representations (ICLR). Vancouver, Canada, 2018: 1-16.
- [7] Geng X, Li Y, Wang L, et al. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting// Proceedings of the 33th AAAI Conference on Artificial Intelligence (AAAI). Hawaii, USA, 2019: 3656-3663.
- [8] Wu Z, Pan S, Long G, et al. Graph wavenet for deep spatial-temporal graph modeling//Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI). Macao, China, 2019: 1907-1913
- [9] Huang S, Wang D, Wu X, et al. Dsanet: Dual self-attention network for multivariate time series forecasting//Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM). Beijing, China, 2019: 2129-2132.
- [10] Lv Z, Xu J, Zheng K, et al. LC-RNN: A deep learning model for traffic speed prediction//Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI). Stockholm, Sweden, 2018: 3470-3476.
- [11] Song J, Son J, Seo DH, et al. ST-GAT: A Spatio-temporal graph attention network for accurate traffic speed prediction// Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM). Atlanta, USA, 2022: 4500-4504.
- [12] Qin H, Ke S, Yang X, et al. Robust spatio-temporal purchase prediction via deep meta learning//Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI). 2021: 151-159.
- [13] Wang, F, Xu, J, Liu, C, et al. On prediction of traffic flows in smart cities: a multitask deep learning based approach// Proceedings of the 30th International Conference on World Wide Web (WWW). Ljubljana, Slovenia, 2021: 805-823.
- [14] Rossi E, Chamberlain B, Frasca F, et al. Temporal graph networks for deep learning on dynamic graphs. ICML 2020 Workshop on Graph Representation Learning, 2020: 1-16.
- [15] Xu D, Ruan C, Korpeoglu E, et al. Inductive representation learning on temporal graphs//Proceedings of the 8th International Conference on Learning Representations (ICLR). Addis Ababa, Ethiopia, 2020: 1-19.
- [16] Li F, Feng J, Yan H, et al. Dynamic graph convolutional recurrent network for traffic prediction; Benchmark and solution. ACM Transactions on Knowledge Discovery from Data (TKDD). 2022, 17(1): 1-13.
- [17] Bai L, Yao L, Li C, et al. Adaptive graph convolutional recurrent network for traffic forecasting//Proceedings of the 33rd Advances in Neural Information Processing Systems (Neurips). 2020: 17804-17815.
- [18] Schölkopf B, Locatello F, Bauer S, et al. Toward causal representation learning. Proceedings of the IEEE. 2021, 109(5): 612-634.
- [19] Varian H R. Causal inference in economics and marketing. Proceedings of the National Academy of Sciences (PNAS). 2016, 113(27): 7310-7315.
- [20] Kaddour J, Zhu Y, Liu Q, et al. Causal effect inference for structured treatments//Proceedings of the 34th Advances in Neural Information Processing Systems (Neurips). 2021: 24841-24854.
- [21] Gangl M. Causal inference in sociological research. Annual review of sociology, 2010, 36(1): 21-47.
- [22] Lin X, Chen Y, Li G, et al. A Causal Inference Look At Unsupervised Video Anomaly Detection//Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). 2022: 1620-1629.
- [23] Liu Y, Cadei R, Schweizer J, et al. Towards robust and adaptive motion forecasting: a causal representation perspective//Proceedings of the 35th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, America, 2022: 17081-17092.
- [24] Yang M, Liu F, Chen Z, et al. CausalVAE: Disentangled representation learning via neural structural causal models// Proceedings of the 34th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 9593-9602.
- [25] Yue Z, Zhang H, Sun Q, et al. Interventional few-shot learning//Proceedings of the 33rd Advances in Neural Information Processing Systems (Neurips). 2020: 2734-2746.
- [26] Glymour M, Pearl J, Jewell N P. Causal inference in statistics: A primer. USA: John Wiley & Sons, 2016.
- [27] Kingma D P, Ba J. Adam: A method for stochastic optimization//Proceedings of the 3rd International Conference on Learning Representations (ICLR). San Diego, USA, 2015: 1-15.
- [28] Guo K, Hu Y, Sun Y, et al. Hierarchical graph convolution network for traffic forecasting//Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI). 2021: 151-159.
- [29] Han L, Du B, Sun L, et al. Dynamic and multi-faceted spatio-temporal deep learning for traffic speed forecasting// Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD). 2021: 547-555.



DENG Pan, Ph. D., associate professor. Her research interests include spatio-temporal data mining, traffic flow prediction and causal learning.

LIU Jun-Ting, Ph. D. candidate. His research interests include spatio-temporal data mining, traffic flow prediction and causal learning.

WANG Xiao, Ph. D., professor. His research interests include social computing, parallel vehicle networking, and cognitive automatic driving.

Background

The research in this work belongs to spatiotemporal data mining. The spatiotemporal representation learning task aims to transform unstructured high-dimensional spatiotemporal data into structured low-dimensional feature representation, so that it can effectively represent the potential rules and knowledge in spatiotemporal data, and improve the performance of downstream tasks. Spatiotemporal representation learning is a research hotspot in the field of spatiotemporal data mining, which is widely used in intelligent transportation, urban planning, environmental protection and other fields. Most spatiotemporal representation learning methods usually model the spatial attributes in the spatiotemporal data as a graph structure, and the temporal attributes as a time series on the nodes. The spatiotemporal correlation in the observation data is modeled through recurrent neural network and graph convolution neural network, and feature representation is extracted and used for downstream tasks such as anomaly detection, spatiotemporal sequence prediction, etc. However,

JIA Xiao-Feng, Ph. D., senior engineer. His research interests include complex system optimization and spatiotemporal data mining.

ZHAO Yu, Ph. D. candidate. His research interests include spatio-temporal data mining, traffic flow prediction and causal learning.

WANG Mu-Lan, M. S. candidate. His research interests include spatio-temporal data mining, traffic flow prediction and causal learning.

DAI Xing-Yuan, Ph. D., assistant researcher. His research interests include artificial intelligence, intelligent transportation system and cognitive automatic driving.

due to the imbalance of the datasets and the limitations of the model, the model may learn the spurious spatiotemporal correlations in the observation data. On the one hand, the confounders in the temporal domain lead to non-causal spurious correlation between historical data and future data, which makes the modeling process of the model affected by confounding bias in the temporal domain; On the other hand, the graph convolution networks lead to the non-causal correlation between nodes, which makes the model affected by the spurious spatial correlation between nodes in the feature extraction process, and ultimately leads to generating biased spatiotemporal representation, which greatly degrades the performance of downstream tasks. In this work, we present a novel framework named Spatial-Temporal Causal Transition Network (STCTN), which eliminates confounding bias in temporal domain and models causal transition process in spatial domain. The proposed model effectively improves the performance of downstream prediction tasks and show its robustness on nodes with different types.