基于结构因果模型的城市出行流量预测方法

刘钰婷，顾晶晶，周强

引用本文

刘钰婷，顾晶晶，周强. 基于结构因果模型的城市出行流量预测方法[J]. 计算机科学，2025，52(10)：70-78.

LIU Yuting, GU Jingjing, ZHOU Qiang. Urban Flow Prediction Method Based on Structural Causal Model [J]. Computer Science, 2025, 52(10): 70-78.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.241000088

基于结构因果模型的城市出行流量预测方法

刘钰婷　顾晶晶　周　强
南京航空航天大学计算机科学与技术学院　南京 210000
(yuting_liu@nuaa.edu.cn)

摘　要　城市出行流量预测是智慧城市研究中的重要课题,为城市规划和资源优化提供了关键的数据支持。近年来,基于图神经网络的城市流量预测模型在提升预测精度上取得了显著进展。然而,大多数现有研究都假设训练数据和测试数据来自相同的分布,忽视了现实世界中城市流量分布动态变化的复杂性,导致模型在面对分布偏移时表现不佳。为了解决这一问题,提出一种基于结构因果模型的城市出行流量预测方法,旨在应对分布偏移带来的模型泛化挑战。该方法首先利用结构因果模型揭示环境因素作为混淆变量对流量预测的影响效应,并设计共享分布估计器以学习环境信息的先验分布,进而引入后门调整方法,结合变分推断有效消除环境因素引起的混淆影响。该方法能够公平地考虑不同环境信息,提升流量预测的准确性与鲁棒性。在两个真实世界数据集上的实验结果表明,所提方法在应对分布偏移时具有较高的预测精度和鲁棒性。与6种主流基线模型相比,预测性能提升了2.26%～9.18%。

关键词:城市出行流量预测;因果推断;分布偏移;时空数据挖掘;结构因果模型
中图分类号　TP183

Urban Flow Prediction Method Based on Structural Causal Model

LIU Yuting,GU Jingjing and ZHOU Qiang
College of Computer Science and Technology,Nanjing University of Aeronautics and Astronautics,Nanjing 210000,China

Abstract　Urban flow prediction plays a critical role in smart city research,providing essential data for urban planning and resource optimization. In recent years,Graph Neural Network (GNN)-based models have significantly enhanced the accuracy of urban flow prediction. However,most existing studies assume that training and testing data come from the same distribution,ignoring the complexity of dynamic changes in urban flow distribution in the real world,leading to poor model performance. To address this challenge,this paper proposes an urban flow prediction method based on the structural causal model to effectively deal with the challenge of model generalization caused by distribution shift. This method first utilizes a structural causal model to uncover the impact of environmental factors as confounders on flow prediction. It then designs a shared distribution estimator to learn the prior distribution of environmental information. Furthermore,a backdoor adjustment approach is introduced,combined with variational inference,to effectively eliminate the confounding effects caused by environmental factors. The proposed method can fairly consider different environmental factors,improving the accuracy and robustness of prediction. Experimental results on two real-world datasets show that the proposed model has high prediction accuracy and robustness when dealing with distribution shift. Compared with the six state-of-the-art baselines,the prediction performance is improved by 2.26%～9.18%.

Keywords　Urban flow prediction,Causal inference,Distribution shift,Spatio-temporal data mining,Structural causal model

1　引言

随着城市化进程的加速,城市出行流量预测在优化资源配置和提升城市管理效率方面的重要性日益凸显。得益于数据采集技术的进步,传感器、社交媒体以及环境变化等多源数据为流量预测提供了有力的支持。然而,随着时间跨度的延长,这些数据的规模不仅呈现指数级增长,其分布和时空特性也变得愈加复杂。现有的城市流量预测研究[1-3]主要依赖图神经网络来捕捉时空规律。这些模型通常通过预定义或自适应的图结构学习城市节点之间的空间依赖性,并结合时间卷积网络或递归神经网络建模时间依赖。然而,大多数方法[4-6]都假设训练和测试数据来自相同的分布,而这种假设在现实城市流量场景中往往不成立。随着时间推移,城市流量往往呈现出多样化且动态变化的分布特征。因此,在大规模、复杂且存在分布偏移的场景中实现有效预测成为一个巨大的挑战。

到稿日期:2024-10-17　返修日期:2025-02-25
基金项目:国家自然科学基金面上项目(62072235);江苏省自然科学基金青年项目(BK20241402)
This work was supported by the Natural Science Foundation of China(62072235) and Young Scientists Fund of the Natural Science Foundation of Jiangsu Province,China(BK20241402).
通信作者:顾晶晶(gujingjing@nuaa.edu.cn)

城市流量的观测数据往往高度依赖于环境因素。例如，晴天时公园流量可能维持在日常水平；当暴风雪等恶劣天气发生时，公园的人流量显著减少。而在音乐节期间，由于活动的吸引力，公园吸引了大批游客，出行流量激增，远超平日水平。这表明不同环境会显著影响人们的出行决策，导致出行流量观测数据的巨大差异。这种高度动态的环境因素加剧了城市流量预测的复杂性和不确定性。

解决城市流量中分布偏移导致的难以泛化问题面临几个挑战。首先，随着时间推移，模型需要具备在不同分布中的泛化能力，即能够从训练阶段的已知环境外推到全新的、未见过的环境。然而，大多数现有的方法聚焦于在训练数据上学习模型，并基于同一分布的数据进行评估，在应对不同分布测试数据时，往往表现出次优的泛化能力。其次，为了探究环境因素的影响，需要了解环境因素的潜在分布。然而，环境因素（如极端天气、交通事故和大型活动等）具有高度的多样性和难以预测性，这显著增加了数据收集与建模的难度。全面覆盖所有可能的环境因素并准确获取其先验分布几乎是不现实的。因此，全面理解并精确建模环境变化对出行流量的动态影响十分困难。

为了应对上述两个挑战，本文提出一种基于结构因果模型的城市出行流量预测方法，旨在解决城市流量预测中分布偏移导致的模型泛化困难的问题。具体来说，首先从数据生成机制出发，利用结构因果模型对环境因素、历史观测序列和未来流量之间的因果关系进行分析，证明环境因素在数据生成过程中充当了混淆变量，导致模型在训练时依赖虚假相关性，从而难以泛化到新的数据分布或场景中，这也解释了现有模型在面对不同分布的数据时表现不佳的原因。针对这一问题，本文设计了一个分布估计器，通过共享结构，结合真实流量序列和伪流量序列来学习环境因素的先验和变分后验，从而实现后门调整以消除混淆影响。这种基于因果干预的设计为城市流量预测提供了更加稳健的解决方案，能够更准确地应对复杂动态环境下的城市流量变化。

本文的主要贡献概括如下：

1）提出了一种基于结构因果模型的城市出行流量预测方法 CauDS，旨在有效解决城市流量预测中分布偏移导致难以泛化的问题，系统分析了环境因素对数据生成过程的影响，并通过因果干预增强了模型在不同分布条件下的泛化能力。

2）设计了一个共享分布估计器来学习环境因素的先验分布和变分分布，其无需依赖对大规模环境数据的获取，大幅提升了计算效率。

3）在两个真实世界数据集上的广泛实验表明，CauDS 能够显著缓解分布偏移问题，在城市出行流量预测任务中表现出了更高的精度和鲁棒性。

2 相关工作

2.1 城市流量预测

智慧城市的蓬勃发展使得城市流量预测成为其中一个至关重要的研究课题。传统的基于时空图神经网络的模型，例如 STGCN[7] 和 STSGCN[8]，通常基于预定义的图结构来捕获节点之间的空间依赖，同时结合时间卷积网络或递归神经网络建模时间依赖。然而，这类模型在处理复杂、多变的城市流量时存在一定的局限性，主要是因为它们依赖于固定的图结构，难以灵活地捕获节点之间潜在的、隐藏的依赖关系。为了突破这些限制，基于自适应图结构的方法应运而生。例如，GWNET[9] 和 AGCRN[10] 引入可学习的自适应邻接矩阵，使模型能够动态调整节点之间的关系，不再局限于预定义的固定结构。这种自适应机制使得模型能够应对城市流量中的复杂变化和不确定性。此外，一些工作进一步引入注意力机制来增强模型的预测能力。例如，GMAN[4] 通过时间注意力和空间注意力分别建模时间和空间特征，显著提升了长时预测的能力。ASTGCN[11] 通过时空注意力使网络聚焦于关键的时空信息，并结合时空卷积模块进一步捕捉时空依赖。

尽管这些方法在时空依赖性建模方面取得了显著进展，但它们普遍假设训练数据与测试数据来自相同分布。这种假设导致它们在处理广泛存在的城市分布偏移场景时表现出明显的不足，无法有效应对分布偏移带来的泛化困难问题。

2.2 因果推断

因果推断的主要目标是识别和量化变量之间的因果关系，而不仅仅是揭示它们之间的相关性。近年来，因果推断理论与深度学习技术的结合受到了广泛关注，现有研究[12-14] 显示，其具有巨大的潜力。Zhang 等[15] 利用结构因果模型分析弱监督语义分割模型中图像、上下文和类标签之间的因果关系，并将伪掩码边界模糊归结为上下文混淆，提出使用因果干预切断上下文和图像之间的关联，以提升伪掩码质量。Niu 等[16] 提出了一个反事实推理框架，将语言偏见建模为问题对答案的直接因果效应，通过从总因果效应中减去直接语言效应来减少语言偏见。Ge 等[17] 设计了一种社会交叉注意力机制，用于对轨迹特征进行因果干预，以消除历史和未来轨迹之间的虚假相关性。基于此背景，本文也借助因果推断这一强大的工具来解决分布外泛化问题，以提高预测的准确性和鲁棒性。

3 问题定义

本章首先对相关概念进行定义，并正式阐述城市流量预测问题。表 1 总结了本文使用的主要数学符号。随后，将结合城市出行流量预测场景介绍相关因果概念。

表 1 数学符号定义

Table 1 Definitions of mathematical symbols

符号	描述
X, X_t	历史 T 个时间和第 t 个时间的城市流量
$Y, \hat{Y}$	未来 S 个时间的城市流量和预测值
K, D'	环境因素的数量和表征维度
N	城市中区域（节点）的数量
D	城市流量的特征数量
S, S'	样本数量和伪输入数量
X'	流量伪输入
C, c_i	影响城市流量的环境因素和第 i 个环境因素
h_t, h'_{ti}	城市流量的隐藏状态表征
h_t''	因果干预后的城市流量表征
$P(C), P_\lambda(C)$	环境因素的先验分布
$Q_\phi(C\|X)$	环境因素的变分后验分布
$Q_\phi(C\|X')$	伪变分后验分布
$\mathcal{F}(\cdot)$	预测模型 CauDS
$\mathcal{G}(\cdot)$	环境特定的编码器
θ_i, θ_2	环境特定的编码器和预测器的参数
γ_1, γ_2	损失函数各项的贡献

定义 1(城市流量) N 个区域(也称为节点)在时间步 t 上的城市流量值表示为$\boldsymbol{X}_t=\{x_t^1,x_t^2,\cdots,x_t^N\}\in\mathbb{R}^{N\times D}$。其中,$D$ 是观测特征的数量。

定义 2(环境因素) 环境因素指引起城市流量快速波动的外部条件,如天气状况、交通事故、节假日以及特殊事件(节庆活动、体育赛事等)等。在本文中,环境因素 $\boldsymbol{C}$ 被定义为一组可学习的离散变量,表示为 $\boldsymbol{C}=\{\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_K\}\in\mathbb{R}^{K\times D'}$,$K$ 表示环境因素的数量,D'表示特征的维度。

问题定义:基于分布偏移数据的城市出行流量预测问题。给定过去 T 个时间步的城市流量,目标是学习一个函数 $\mathcal{F}(\cdot)$,该函数用于预测未来 S 个时间步的城市流量值,其中训练集的数据分布不等于测试集的数据分布。具体而言,该预测模型可以表示为:

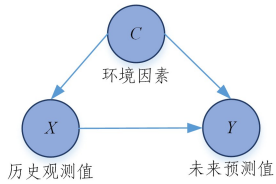
$$[\boldsymbol{X}_{(t-T+1):t}]\xrightarrow{\mathcal{F}(\cdot)}[\boldsymbol{Y}_{(t+1):(t+S)}] \tag{1}$$

为了表达简洁,在后续的章节中使用 $\boldsymbol{X}$ 表示$\boldsymbol{X}_{(t-T+1):t}$,$\boldsymbol{Y}$ 表示$\boldsymbol{Y}_{(t+1):(t+S)}$。

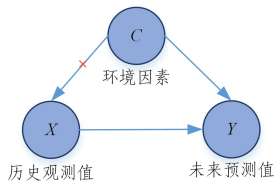
4 基于结构因果模型的城市出行流量预测模型

4.1 基于结构因果模型的因果解释

本节从因果视角深入分析了城市流量与环境因素之间的关系。为了系统化地理解这些因果关系,首先构建了结构因果模型[18](Structural Causal Model, SCM)来解释环境因素 $\boldsymbol{C}$、历史观测序列 $\boldsymbol{X}$ 和未来预测值 $\boldsymbol{Y}$ 之间的因果关系。SCM 结构如图 1(a)所示。在 SCM 中,通过有向图来表示各个因素之间的因果关系,其中有向边的方向表示因果关系流向,即"因"→"果"。



(a)干预前的 SCM



(b)干预后的 SCM

图 1 本文场景下的结构因果模型

Fig. 1 Structural causal model in our scenario

$\boldsymbol{C}\rightarrow\boldsymbol{Y}$:环境因素 $\boldsymbol{C}$ 会影响城市流量预测值 $\boldsymbol{Y}$。环境因素(如天气、突发事件和社会活动等)通过改变人们的出行意愿和行为模式,从而对城市流量产生显著影响。例如,恶劣的天气条件(暴雨、暴雪等)通常会导致公共场所的人流量显著减少,因为不利的出行条件会使人们减少外出或者改变出行计划。而大型活动或突发事件(节日庆典或突发的公共卫生事件)会显著增加或减少特定区域的人流量。因此,在 SCM 中,$\boldsymbol{C}$ 通过直接影响人们的行为模式和流动性,从而影响了人流量的预测值 $\boldsymbol{Y}$。

$\boldsymbol{C}\rightarrow\boldsymbol{X}$:历史城市流量观测值 $\boldsymbol{X}$ 受环境因素 $\boldsymbol{C}$ 影响的原因与 $\boldsymbol{C}\rightarrow\boldsymbol{Y}$ 相同。具体来说,环境因素 $\boldsymbol{C}$ 的存在会显著影响城市流量的时空分布,导致城市流量序列(包括历史观测值和未来预测值)不再遵循正常的模式。在模型中考虑环境因素 $\boldsymbol{C}$,能够有效地识别和调整因这些环境因素引起的异常模式,提高模型的预测精度和鲁棒性,增强对复杂时空数据变化的适应能力。

$\boldsymbol{X}\rightarrow\boldsymbol{Y}$:该条边表示通过历史城市流量观测值 $\boldsymbol{X}$ 推断出未来的城市流量数量 $\boldsymbol{Y}$。该关系是预测模型所要捕捉的核心目标,即将历史数据 $\boldsymbol{X}$ 作为输入,模型通过捕捉时空序列数据中的内在规律,生成准确且鲁棒的未来流量 $\boldsymbol{Y}$ 的预测。

从图 1(a)中可见,存在一条后门路径:$\boldsymbol{X}\leftarrow\boldsymbol{C}\rightarrow\boldsymbol{Y}$。结合上述分析可以明确,环境因素 $\boldsymbol{C}$ 作为混杂因素,导致历史观测值 $\boldsymbol{X}$ 中包含了环境因素的影响,从而破坏了 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 之间的直接因果关系。这种混杂效应会导致模型在训练过程中捕捉到一些不准确或者虚假的关联,进而影响预测的准确性。因此,4.2 节将介绍通过因果干预的方法来消除混杂因素来带的偏差,提高预测的准确性和鲁棒性。

4.2 基于后门准则的因果干预

为了减轻混杂因素 $\boldsymbol{C}$ 的影响,本文采用基于后门调整[19]的因果干预方法,即 $P(\boldsymbol{Y}|\mathrm{do}(\boldsymbol{X}))$。它通过阻断图 1(a)中从 $\boldsymbol{C}$ 到 $\boldsymbol{X}$ 的后门路径$\boldsymbol{X}\leftarrow\boldsymbol{C}\rightarrow\boldsymbol{Y}$(见图 1(b)),有效消除环境因素 $\boldsymbol{C}$ 引起的混淆效应。具体来说,后门调整公式如下:

$$P(\boldsymbol{Y}|\mathrm{do}(\boldsymbol{X}))=\sum_{i=1}^{K}P(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{C}=\boldsymbol{c}_i)P(\boldsymbol{C}=\boldsymbol{c}_i) \tag{2}$$

其中,$P(\boldsymbol{C})$表示环境因素的先验分布,其独立于 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$。通过因果干预方法,$\boldsymbol{C}$ 不再与 $\boldsymbol{X}$ 相关,使得模型在基于 $\boldsymbol{X}$ 预测 $\boldsymbol{Y}$ 时,能够公平地考虑每个环境因素的影响,从而有效避免环境因素引起的混淆偏差。

4.3 学习目标

直接计算式(2)是一项极具挑战性的任务。尽管环境因素 $\boldsymbol{C}$ 是可观测的,但是由于数据收集的局限性,获取全面的环境因素数据在实际操作中是难以实现的,这意味着无法直接获得 $\boldsymbol{C}$ 的先验分布 $P(\boldsymbol{C})$,从而给计算带来了困难。近年来,变分推断[20]得到了广泛的研究和应用[21-22],故本文借鉴变分推断的基本思想,将学习目标式(2)转换为一个优化问题:

$$\log P_\theta(\boldsymbol{Y}|\mathrm{do}(\boldsymbol{X}=\boldsymbol{x}))\geqslant\mathbb{E}_{\boldsymbol{c}\sim Q_\phi(\boldsymbol{C}|\boldsymbol{X}=\boldsymbol{x})}[\log P_\theta(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{C}=\boldsymbol{c}_i)]-KL)Q_\phi(\boldsymbol{C}|\boldsymbol{X})\,\|\,P_\lambda(\boldsymbol{C})) \tag{3}$$

其中,$Q_\phi(\boldsymbol{C}|\boldsymbol{X})$是变分分布,$P_\lambda(\boldsymbol{C})$是先验分布。式(3)的推导基于 Jensen 不等式,等号仅在变分分布$Q_\phi(\boldsymbol{C}|\boldsymbol{X})$与先验分布 $P_\lambda(\boldsymbol{C})$相等时成立。式(3)的第一项是预测误差;第二项是 Kullback-Leibler(KL)散度,它鼓励变分分布尽可能地接近环境因素的先验分布。

4.4 模型结构

本文提出了基于结构因果模型的城市流量出行预测模型 CauDS,旨在有效解决城市流量出行预测中分布偏移导致的泛化困难问题。模型主要由共享分布估计器、特定环境编码器和流量预测器 3 个部分组成,整体架构如图 2 所示。

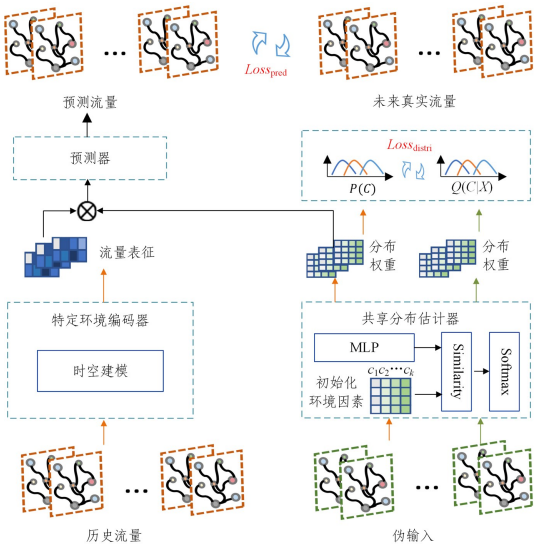


图 2 CauDS 总体框架图

Fig. 2 Framework overview of CauDS

4.4.1 共享分布估计器

本小节提出一个共享分布估计器，用于估计环境因素的变分分布$Q_\phi(\boldsymbol{C}|\boldsymbol{X})$和先验分布$P_\lambda(\boldsymbol{C})$。

给定第 t 个时间戳的城市流量序列$\boldsymbol{x}_t \in \mathbb{R}^D$，首先将其投影到与环境因素 $\boldsymbol{C}$ 相同的空间，以计算流量序列中所包含的环境信息。然后应用 Softmax 函数获得城市流量序列在第 i 个环境因素下的分布概率，表示为：

$$Q_\phi^{t,i}=\frac{\exp((\boldsymbol{x}_t\boldsymbol{W}\cdot\boldsymbol{c}_i)/\tau)}{\sum_j^K\exp((\boldsymbol{x}_t\boldsymbol{W}\cdot\boldsymbol{c}_j)/\tau)} \tag{4}$$

其中，$\boldsymbol{W}$ 为投影矩阵，τ 是温度系数。进一步，使用 $Estimator(\boldsymbol{X})$ 表示城市流量序列 $\boldsymbol{X}$ 的分布估计：

$$Estimator(\boldsymbol{W})=[Q_\phi^{t,i}]_{i=\{1,\cdots,K\}}^{t=\{1,\cdots,T\}}\in\mathbb{R}^{T\times K} \tag{5}$$

1）变分分布$Q_\phi(\boldsymbol{C}|\boldsymbol{X})$估计：将流量数据输入共享分布估计器中，以拟合变分分布，表示为$Q_\phi(\boldsymbol{C}|\boldsymbol{X})=Estimator(\boldsymbol{W})$。

2）先验分布$P_\lambda(\boldsymbol{C})$估计：现有的大部分方法采用预先定义的均匀分布来近似先验[23]，但这种简单的分布难以捕捉真实世界中复杂多变的环境因素分布。因此，已有方法[24]通过聚合后验来估计先验，以应对复杂的现实世界。

$$P_\lambda(\boldsymbol{C})=\frac{1}{S}\sum_{n=1}^{S}Q_\phi(\boldsymbol{C}|\boldsymbol{X}) \tag{6}$$

其中，S 是数据集中的样本总数。然而，该方法使用了所有的训练数据，容易导致过拟合，并且其计算成本非常高。为了解决上述问题，受现有研究[25]使用高斯混合模型作为灵活且可学习先验的启发，本文使用伪变分后验混合作为环境因素先验分布的估计。具体而言，通过随机生成的伪序列 $\boldsymbol{X}'$来模拟数据，伪序列的数量S'远小于真实样本数量 S，先验分布$P_\lambda(\boldsymbol{C})$可以表示为：

$$P_\lambda(\boldsymbol{C})=\frac{1}{S'}\sum_{n=1}^{S'}Q_\phi(\boldsymbol{C}|\boldsymbol{X}') \tag{7}$$

其中，S'表示伪输入的数量，其远小于样本数量 S；$Q_\phi(\boldsymbol{C}|\boldsymbol{X}')=Estimator(\boldsymbol{X}')$。这些伪输入随机生成并在训练过程中通过反向传播学习，可以视为先验分布的超参数。

该估计器不仅能够有效估计环境因素分布，还通过避免对大量环境因素进行繁琐检索，显著提高了计算效率。利用共享分布估计器，可以在更短的时间内获得可靠的估计结果，使其在处理大规模数据集或高维时空数据中尤其具有优势。

4.4.2 环境特定的编码器

为了有效捕捉和学习不同环境下的时空序列特征，即后门调整公式的第一项 $P(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{C}=\boldsymbol{c}_i)$，本文提出了环境特定的编码器$\mathcal{G}(\cdot)$。该编码器的核心思想是，为每个环境因素引入一组独立的可训练参数，以适应各自环境中的独特特征和模式。在模型训练过程中，编码器能够基于环境信息动态调整参数配置，从而灵活应对不同环境条件。

t 时刻在第 i 个环境因素c_i下的隐藏状态可以表示为：

$$\boldsymbol{h}'_{ti}=\mathcal{G}(\boldsymbol{h}_t,\boldsymbol{c}_i;\boldsymbol{\theta}_i) \tag{8}$$

其中，$\boldsymbol{h}_t$和$\boldsymbol{c}_i$分别表示当前时间戳的序列隐藏状态和环境因素表征，$\boldsymbol{h}'_{ti}$表示下一时间戳的序列隐藏状态，$\boldsymbol{\theta}_i$是模型参数。编码器$\mathcal{G}(\cdot)$可为任意的时空建模模型，本文使用经典的时空建模模型 STGCN[1]作为基准实现。

通过引入环境特定的编码器增强了模型在复杂、多变的时空环境下的特征提取能力，也显著提升了其预测性能。

4.4.3 因果干预的实例化

为了实例化完整的式(2)，以实现基于后门调整的因果干预，将得到的环境分布概率$Q_\phi^{t,i}$与对应的时空表征结合，得到第 t 个时间步经过因果干预后的表征：

$$\boldsymbol{h}_t''=\sum_{i=1}^{K}Q_\phi^{t,i}\cdot\boldsymbol{h}''_{ti} \tag{9}$$

其中，$Q_\phi^{t,i}$是第 i 个环境因素的概率，$\boldsymbol{h}''_{ti}$表示第 t 个时间戳在环境因素 $\boldsymbol{c}_i$下的编码器状态。经过因果干预后的表征公平地考虑了每个环境因素的影响，反映了序列隐藏状态在不同环境因素下的综合表征。最终，输出可以表示为：

$$\boldsymbol{H}=[\boldsymbol{h}''_{t-T+1},\cdots,\boldsymbol{h}''_{t-1},\boldsymbol{h}_t'']\in\mathbb{R}^{T\times D} \tag{10}$$

4.4.4 流量预测器

将综合特征向量 $\boldsymbol{H}$ 输入预测器中，生成最后的城市流量预测结果$\hat{\boldsymbol{Y}}$：

$$\hat{\boldsymbol{Y}}=Predictor(\boldsymbol{H};\boldsymbol{\theta}_2) \tag{11}$$

其中，$Predictor$ 是一个两层的感知机，$\boldsymbol{\theta}_2$是对应的参数。为了衡量预测结果的准确性，本文使用平均绝对误差 MAE 作为预测损失函数，定义如下：

$$L_{\text{pred}}=\frac{1}{S}\sum_{i=1}^{S}|\boldsymbol{Y}_i-\hat{\boldsymbol{Y}}_i| \tag{12}$$

其中，S 是样本总数。

在训练过程中，CauDS 在下述损失函数监督下学习：

$$L=\gamma_1 L_{\text{dictri}}+\gamma_2 L_{\text{pred}} \tag{13}$$

其中，$L_{\text{dictri}}=KL(Q_\phi(\boldsymbol{C}|\boldsymbol{X})\parallel P_\lambda(\boldsymbol{C}))$是变分分布$Q_\phi(\boldsymbol{C}|\boldsymbol{X})$和先验分布$P_\lambda(\boldsymbol{C})$之间的 KL 散度；参数$\gamma_1$和$\gamma_2$分别用于控制$L_{\text{dictri}}$和$L_{\text{pred}}$在总损失中的贡献，确保模型在不同目标下均衡学习。

5 实验与结果分析

5.1 数据集

本文在纽约出租车数据集和北京地图查询轨迹数据集上

进行实验，来验证所提模型的有效性。其中，纽约数据集来自完全公开的真实数据集，北京数据集来自第三方位置服务供应商提供的地图查询轨迹数据集。表 2 总结了北京和纽约城市流量数据集的统计信息。

表 2 城市流量数据集的统计信息

Table 2 Statistics of urban flow datasets

数据集描述	纽约	北京
时间跨度	2012.06－2016.12	2018.07－2019.10
记录数量	35280	10241
节点数量	160	185
记录最小粒度	小时	小时

5.2 对比方法和评价指标

为了全面评估 CauDS 在处理城市流量预测任务中的性能和优势，本文将 CauDS 模型与以下 6 种经典和最新的基线进行比较。

1）HA：一种传统的线性流量预测方法，通过对历史同一时间段的数据求平均来预测未来的流量值。

2）STGCN[7]：一种时空图卷积网络，它在图结构上进行建模并采用卷积结构，能有效地捕捉时空依赖关系。

3）GWNet[9]：提出了一种通过节点嵌入学习自适应依赖矩阵的机制，使得模型可以自动捕获数据中隐藏的空间依赖关系。

4）STSGCN[8]：通过构造局部时空图并应用时空同步建模机制，能够有效地捕捉局部时空图中复杂的时空相关性。

5）AGCRN[10]：将图卷积和 GRU 结合，通过节点自适应图卷积捕获节点特定模式，使得模型能够灵活适应不同节点的特性和模式。

6）CauSTG[26]：一个时空因果学习框架，旨在将不变关系迁移到分布外场景。其通过时空一致性学习器和分层不变性探索器，能够提取时空数据中的不变关系。

本文采用平均绝对误差（MAE）、均方根误差（RMSE）和平均绝对百分比误差（MAPE）来评估模型的性能。MAE，RMSE 和 MAPE 的计算如式(14)－式(16)所示：

$$MAE=\frac{1}{S}\sum_{i=1}^{S}|\hat{\boldsymbol{Y}}_i-\boldsymbol{Y}_i| \tag{14}$$

$$RMSE=\sqrt{\frac{1}{S}\sum_{i=1}^{S}(\hat{\boldsymbol{Y}}_i-\boldsymbol{Y}_i)^2} \tag{15}$$

$$MAPE=\frac{100\%}{S}\sum_{i=1}^{S}\left|\frac{\hat{\boldsymbol{Y}}_i-\boldsymbol{Y}_i}{\boldsymbol{Y}_i}\right| \tag{16}$$

其中，S 表示样本总数，$\hat{\boldsymbol{Y}}_i$ 表示城市流量预测值，$\boldsymbol{Y}_i$ 表示城市流量真实值。MAE，RMSE 和 MAPE 3 个评价指标值均越小越好。

5.3 实验设置

实验按照 7∶1∶2 的比例划分训练集、验证集和测试集，用过去 $T=12$ 个时间步（12 小时）的历史交通流量去预测未来 $S=[6,12]$（6 小时和 12 小时）的交通流量。纽约和北京数据集的观测特征数量 D 均为 1。在纽约数据集上，环境因素的数量 K 设置为 5；在北京数据集上，K 设置为 4。伪输入 S' 均设置为 $2N$，N 为数据集中的节点数量。流量和伪输入的隐藏表征维度均为 64。预测器 Predictor 是 2 层多层感知机。本文方法和所有的基线都是用 PyTorch 实现的。模型使用学习率为 0.001 的 Adam 优化器进行训练，批大小设置为 64。

5.4 实验结果分析

5.4.1 整体性能比较

为了验证本文模型的有效性，将本文模型与现在基线模型进行比较，评估了它们在纽约和北京数据集上关于 MAE，RMSE 和 MAPE 的表现。最佳结果用粗体突出显示，次优结果用下划线显示。纽约数据集上的实验结果如表 3 所列，北京数据集上的实验结果如表 4 所列。

从表 3 和表 4 的实验结果可以看出，CauDS 在纽约和北京这两个数据集上的表现均明显优于其他基线模型，证明了该模型的有效性和泛化能力。这些结果表明，提出的因果驱动方法能够有效应对分布偏移问题，特别是在复杂的时空数据环境中展现出了出色的预测能力。此外，最简单的基线模型 HA 表现最差。这是因为 HA 只基于历史数据的平均值来进行预测，无法捕捉数据中的复杂时空动态变化，难以应对复杂的流量波动场景，所以预测性能较低。相比之下，基于预定义图结构的图神经网络模型（如 STGCN，STSGCN）在实验中表现出了更好的效果。由于它们能够捕捉到数据中的时空依赖关系，通过图结构建模城市流量的空间依赖，因此显著提升了预测的精度。带有自适应邻接矩阵的图神经网络模型（如 GWNet，AGCRN）的性能超越了使用预定义图结构模型的性能，因为自适应邻接矩阵可以根据数据的特性，动态地学习节点之间的隐藏依赖关系，避免了预定义图结构中可能存在的信息丢失或依赖不充分的问题。CauSTG 在所有基线模型中基本取得了最佳效果，仅在少数评估指标上略逊色于 AGCRN，这是由于 CauSTG 引入了时空一致性学习器和分层不变性探索器，通过学习时空观测中的不变性来提升模型的鲁棒性。在面对分布偏移的情况下，CauSTG 能够更好地保证模型的泛化能力，展现了其在应对复杂城市流量预测任务时的卓越表现。

表 3 纽约数据集上的模型性能比较

Table 3 Model performance comparison on NYC dataset

Method	6h			12h			平均		
	MAE	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*
HA	102.47	246.88	0.85	102.47	246.88	0.85	102.47	246.88	0.85
STGCN	57.58	148.26	0.56	59.05	156.94	0.57	58.32	152.60	0.57
GWNet	42.35	101.31	0.39	45.80	102.34	0.40	44.08	101.83	0.40
STSGCN	46.99	113.92	0.39	49.71	115.71	0.41	48.35	114.82	0.40
AGCRN	40.74	<u>92.09</u>	0.38	43.19	95.87	0.39	41.97	93.98	0.39
CauSTG	<u>39.73</u>	93.31	<u>0.31</u>	<u>39.85</u>	<u>90.46</u>	<u>0.33</u>	<u>39.79</u>	<u>91.89</u>	<u>0.32</u>
CauDS	**36.64**	**88.89**	**0.30**	**36.50**	**89.61**	**0.31**	**36.57**	**89.25**	**0.30**

表 4 北京数据集上的模型性能比较

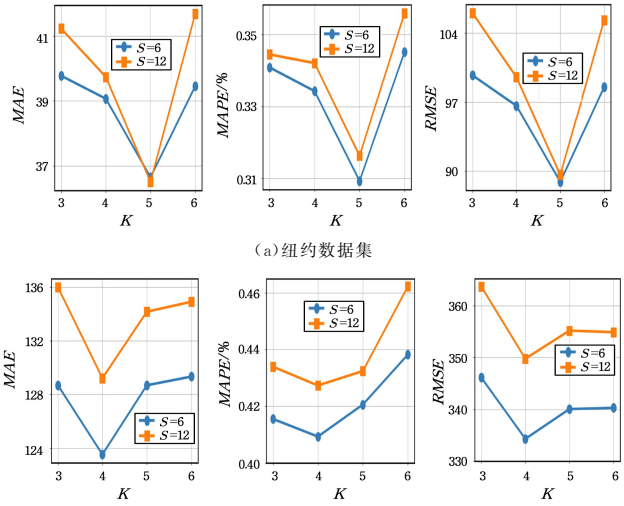
Table 4 Model performance comparison on Beijing dataset

Method	6 h			12h			平均		
	MAE	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*	*MAE*	*RMSE*	*MAPE*
HA	289.91	756.09	0.81	289.91	756.09	0.81	289.91	756.09	0.81
STGCN	176.84	565.16	0.45	181.53	570.63	0.46	179.19	567.90	0.46
GWNet	136.13	367.09	0.45	138.12	386.20	0.46	137.13	376.65	0.46
STSGCN	176.28	527.05	0.45	194.12	588.14	0.45	185.20	557.60	0.45
AGCRN	128.73	344.32	0.44	137.58	366.91	0.44	133.16	355.62	0.44
CauSTG	126.18	341.88	0.43	134.35	360.35	0.45	130.27	351.12	0.44
CauDS	**123.52**	**334.30**	**0.40**	**129.20**	**349.77**	**0.42**	**126.36**	**342.03**	**0.41**

5.4.2 超参数分析

本小节对模型中的两个关键超参数——环境因素数量 K 和伪输入数量 S' 进行了详细分析，以评估它们对模型性能的影响。

首先，针对环境因素的数量 K，将 K 值从 3 调整到 6，观察不同数量的环境因素对模型预测精度的影响。图 3 展示了纽约和北京数据集上的结果。在纽约数据集上，当 $K=5$ 时模型取得最佳结果；在北京数据集上，最佳结果出现在 $K=4$ 时。



(a)纽约数据集

(b)北京数据集

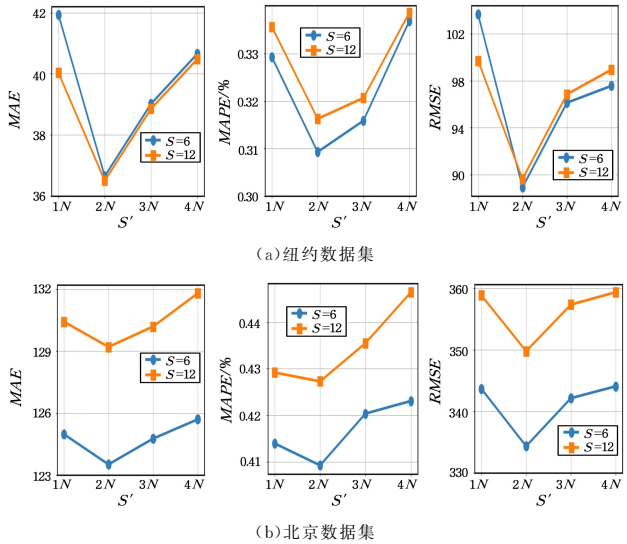
图 3 环境因素数量 K 在纽约和北京数据集上的评估

Fig. 3 Evaluation of the number of environmental factors K on NYC and Beijing datasets

从图 3 中可以看到，随着 K 的增加，模型的性能先是逐步提升，但在达到最佳点后，随着 K 进一步增大，性能反而有所下降。过小的 K 不能有效捕捉复杂的环境特征，导致模型无法充分利用环境信息来提升预测效果；而过大的 K 则可能会引入过多的冗余信息，增加了模型的复杂性，导致过拟合问题的出现，进而影响模型的泛化能力。

接着，改变伪输入的数量 S'，从 $1N$ 到 $4N$，N 表示数据集中节点的数量。图 4 展示了在纽约和北京数据集上的结果。

当 $S'=2N$ 时，在纽约和北京数据集上，模型均达到了最佳性能。这表明，仅需要相对较少数量的伪输入就能够有效学习到环境因素的先验信息，不需要对大量的环境因素进行繁杂的检索和计算。在这种情况下，模型不仅能够维持较高的预测精度，还显著提升了计算效率，避免了大量冗余信息对模型学习过程的干扰。



(a)纽约数据集

(b)北京数据集

图 4 伪输入数量 S' 在纽约和北京数据集上的评估

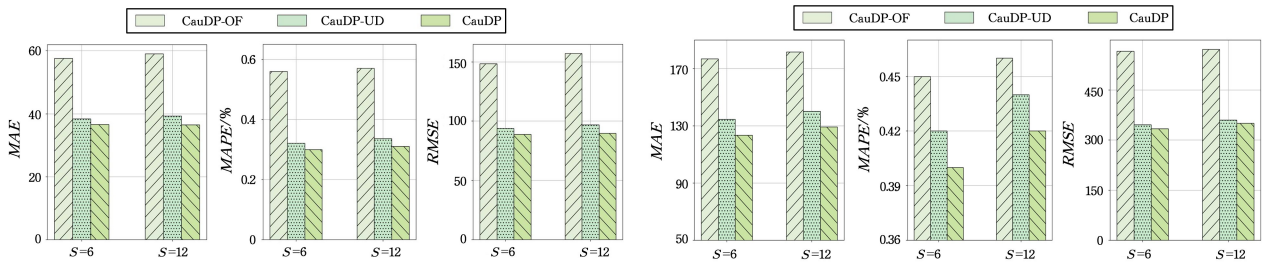
Fig. 4 Evaluation of the number of pseudo input S' on NYC and Beijing datasets

5.4.3 消融实验

本小节对所提 CauDP 模型进行消融实验，用于评估提出的每个模块的有效性。实验包含了以下几种变体。1)CauDP-OF：将环境因素数量设置为 1，此时相当于不考虑环境因素，模型性能取决于基准的选取，本文采用 STGCN 作为基准。2)CauDP-UD：将基于变分推断的先验估计替换为简单的均匀分布。在纽约和北京数据集上的实验结果如图 5 所示。

从图 5 中可得，在两个数据集上，不考虑环境因素的模型 CauDP-OF 的性能急剧下降。这一现象揭示了环境因素在城市流量预测中的重要性。在完全忽略环境因素的情况下，模型无法有效捕捉数据分布的动态变化，尤其是在面对数据分布偏移时，模型的泛化能力受到显著限制。具体来说，当环境条件发生变化时，CauDP-OF 模型未能适应这些变化，导致预测精度大幅下降。

此外，将先验分布估计模块替换为使用均匀分布的模型 CauDP-UD，同样导致了性能的明显下降。这表明，简单地使用均匀分布来替代实际的环境因素分布，无法充分应对现实中的复杂和多变环境。在现实世界中，环境因素并非均匀分布，而是具有复杂的变化模式和非对称性，均匀分布的假设忽视了这些关键特性，因而限制了模型的适应性和泛化能力。相比之下，合理地估计环境因素的分布对于捕捉分布变化、提升预测的鲁棒性具有至关重要的作用。



(a)纽约数据集上的消融实验 (b)北京数据集上的消融实验

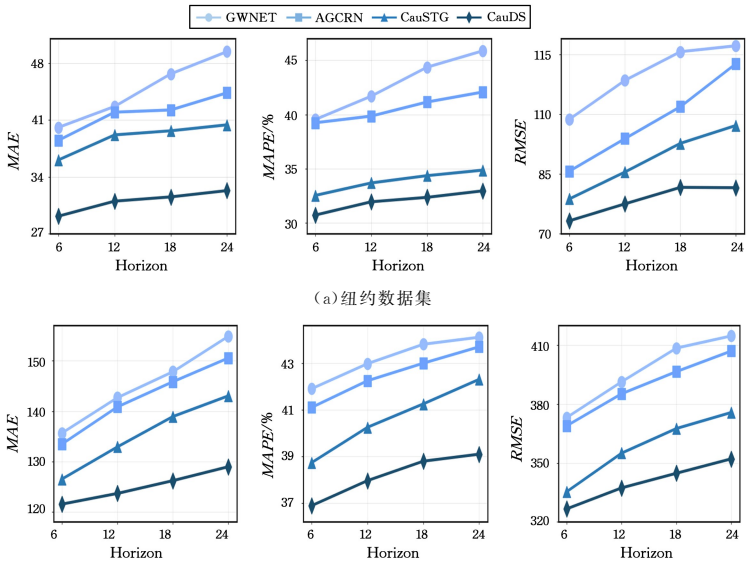
图 5 两个数据集上的消融实验

Fig. 5 Ablation study on NYC and Beijing datasets

5.4.4 不同预测长度的影响

在本节中,对模型的预测长度 S 进行分析,以进一步评估其对模型性能的影响。选取 GWNET,AGCRN 和 CauSTG 3 个基线,将它们与本文提出的 CauDS 在不同预测长度 S 下进行比较,实验结果如图 6 所示。

从图 6 中可知,随着预测长度的增加,所有模型的预测性均有所下降。这是因为模型需要基于历史数据来推断未来的趋势和变化,而历史数据的有限性会对预测精度产生负面影响。然而,CauDS 的预测性能在 S 增大时下降幅度相对较小,表明 CauDS 在长时间步预测任务中具有更强的鲁棒性。其优势可能源于长时间步预测更容易受到显著的分布偏移影响,而 CauDS 可以有效建模由环境因素引起的分布偏移,从而在更长时间步预测中保持优势。此外,未考虑分布偏移的模型 GWNET 和 AGCRN 在预测长度 S 增大时性能下降更为显著,考虑了分布变化的模型 CauSTG 则在长时间步预测中表现更加稳定。这进一步证明了在长时间步预测中建模分布偏移影响的重要性,同时凸显了 CauDS 在此方面的显著优势。



(a)纽约数据集

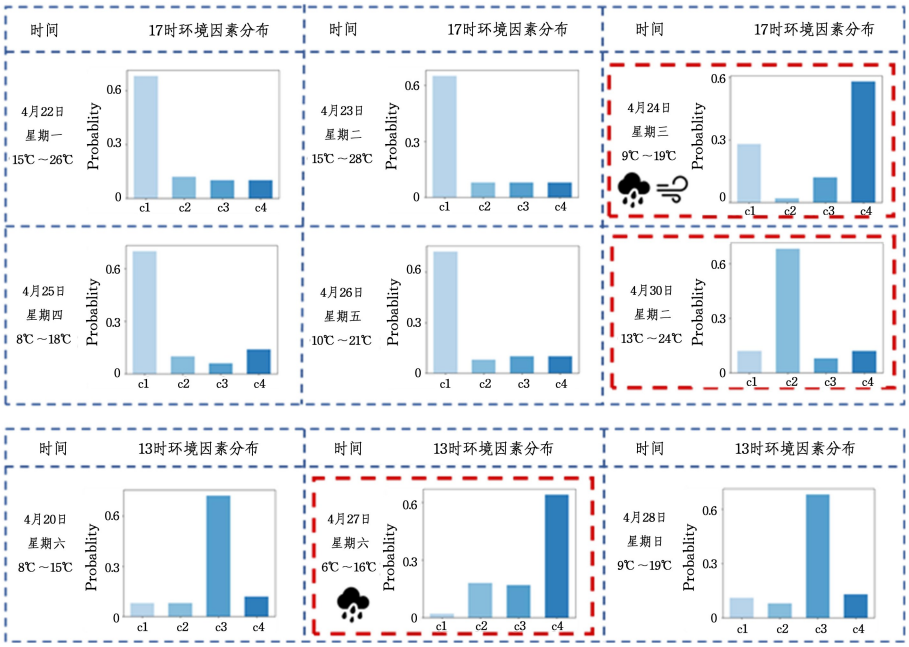
(b)北京数据集

图 6 不同预测长度 S 的实验结果

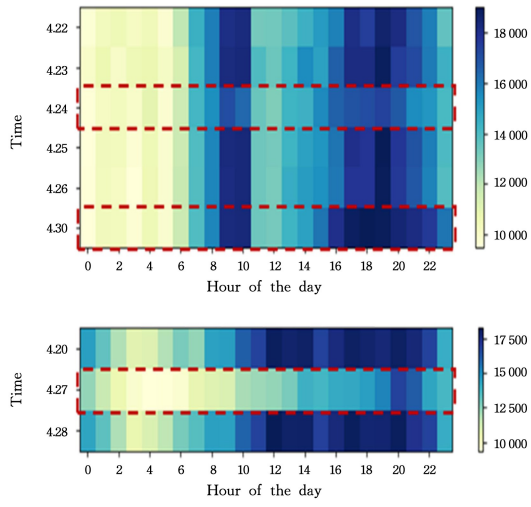
Fig. 6 Experimental results of different prediction lengths S

5.4.5 环境因素分析

为了深入研究环境因素所学习到的信息,图 7 展示了北京数据集 2019 年 4 月 20 日至 4 月 30 日期间某区域的流量分布和学习到的环境因素的分布情况。通过对图中数据的详细分析,可以得到以下几个观察和结论。1)4 月 22 日、4 月 23 日、4 月 25 日和 4 月 26 日的 17 时,它们都有相似的流量分布和环境因素分布,均对应较高的c1环境因素概率,这表明c1可能代表典型的工作日傍晚的环境模式。2)4 月 24 日 17 时,与其他工作日相比,该时段的流量显著降低(见图 7(b))。可以推测,这一变化与当天的降水和大风有关。与此同时,其对应环境因素也与典型的工作日傍晚不同,其c1环境因素概率降低,c4因素概率升高。因此,可以推断c4因素可能代表与恶劣天气有关的环境。3)4 月 30 日 17 时,尽管是工作日傍晚,从图 7(b)中可以看到,它与正常工作日同一时间的流量相比有显著增长,同时,其对应的c2环境因素较高。结合五一小长假前夕的特殊背景,推测大量游客的流入是导致流量增长的原因。因此,c2因素很可能代表小长假模式或者某些特殊事件。4)4 月 20 日和 4 月 28 日的 13 时,它们对应的c3环境因素概率高,这意味着c3可能代表典型的周末中午的环境模式。5)4 月 27 日是星期六,这一天的天气发生了显著变化,出现了降温和降雨。在流量上,这天的流量普遍偏低,明显低于正常周末的流量。然而,这天 13 时,c4因素的概率出现较高,与结论 1)相似。进一步推测,c4大概率代表与恶劣天气或气温状况相关的环境模式。



(a)区域环境因素分布



(b)区域流量分布

图 7 北京数据集上的案例分析

Fig. 7 Case study on Beijing dataset

结束语 本文面向分布偏移导致的模型泛化困难问题，提出了一种基于结构因果模型的城市出行流量预测模型。通过结构因果模型，发现环境因素作为混淆变量对模型的泛化能力有着显著影响。因此，本文设计了一种面向出行流量数据的共享分布估计器，能够合理地学习环境因素的先验分布特征，并通过因果干预消除其引起的混淆效应，从而提升模型在动态分布变化场景下的预测准确性与鲁棒性。在纽约和北京两个真实世界数据集上的实验结果，验证了所提出的方法相比主流的基线方法，具有更高的预测性能。下一步，可以将该方法扩展到更复杂的长期城市流量预测场景中，以进一步提升其在应对更大规模和更复杂分布偏移问题的适应能力。

参 考 文 献

[1] GUO S，LIN Y，GONG L，et al. Self-supervised spatial-temporal bottleneck attentive network for efficient long-term traffic forecasting[C]// 2023 IEEE 39th International Conference on Data Engineering (ICDE). IEEE，2023：1585-1596.

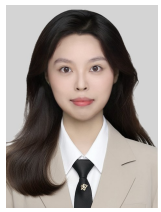
[2] HAN J，ZHANG W，LIU H，et al. BigST：Linear Complexity Spatio-Temporal Graph Neural Network for Traffic Forecasting on Large-Scale Road Networks[C]// Proceedings of the VLDB Endowment. 2024：1081-1090.

[3] PENG H，DU B，LIU M，et al. Dynamic graph convolutional network for long-term traffic flow prediction with reinforcement learning[J]. Information Sciences，2021，578：401-416.

[4] ZHENG C，FAN X，WANG C，et al. Gman：A graph multi-attention network for traffic prediction[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020：1234-1241.

[5] ZHENG Z，GU J，ZHOU Q，et al. Prediction in Long-term Evolution：Exploiting the Interaction Between Urban Crowd Flow Variation and POI Transition Patterns[C]// 2023 IEEE Interna-

tional Conference on Data Mining (ICDM). IEEE, 2023: 1559-1564.
[6] GU J, ZHOU Q, YANG J, et al. Exploiting interpretable patterns for flow prediction in dockless bike sharing systems[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(2): 640-652.
[7] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting[C]// Proceedings of the 27th International Joint Conference on Artificial Intelligence. 2018: 3634-3640.
[8] SONG C, LIN Y, GUO S, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 914-921.
[9] WU Z, PAN S, LONG G, et al. Graph wavenet for deep spatial-temporal graph modeling[C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. 2019: 1907-1913.
[10] BAI L, YAO L, LI C, et al. Adaptive graph convolutional recurrent network for traffic forecasting[J]. Advances in Neural Information Processing Systems, 2020, 33: 17804-17815.
[11] GUO S, LIN Y, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2019: 922-929.
[12] MAGLIACANE S, VAN OMMEN T, CLAASSEN T, et al. Domain adaptation by using causal inference to predict invariant conditional distributions[C]// Advances in Neural Information Processing Systems. 2018.
[13] ZHANG S, YAO D, ZHAO Z, et al. Causerec: Counterfactual user sequence synthesis for sequential recommendation[C]// Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021: 367-377.
[14] ROBERTSM E, STEWART B M, NIELSEN R A. Adjusting for confounding with text matching[J]. American Journal of Political Science, 2020, 64(4): 887-903.
[15] ZHANG D, ZHANG H, TANG J, et al. Causal intervention for weakly-supervised semantic segmentation[J]. Advances in Neural Information Processing Systems, 2020, 33: 655-666.
[16] NIU Y, TANG K, ZHANG H, et al. Counterfactual vqa: A cause-effect look at language bias[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 12700-12710.
[17] GE C, SONG S, HUANG G. Causal intervention for human trajectory prediction with cross attention mechanism[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2023: 658-666.
[18] PEARL J. Models, reasoning and inference [M]. Cambridge: Cambridge University Press, 2000, 19(2): 3.
[19] PEARL J. Causality[M]. Cambridge University Press, 2009.
[20] BLEID M, KUCUKELBIR A, MCAULIFFE J D. Variational inference: A review for statisticians[J]. Journal of the American Statistical Association, 2017, 112(518): 859-877.
[21] DIAOM Z, BALASUBRAMANIAN K, CHEWI S, et al. Forward-backward Gaussian variational inference via JKO in the Bures-Wasserstein space[C]// International Conference on Machine Learning. PMLR, 2023: 7960-7991.
[22] RUDNERT G J, CHEN Z, TEH Y W, et al. Tractable function-space variational inference in bayesian neural networks[J]. Advances in Neural Information Processing Systems, 2022, 35: 22686-22698.
[23] BURDA Y, GROSSE R, SALAKHUTDINOV R. Importance weighted autoencoders[J]. arXiv: 1509.00519, 2015.
[24] HOFFMANM D, JOHNSON M J. Elbo surgery: yet another way to carve up the variational evidence lower bound[C]// Workshop in Advances in Approximate Bayesian Inference, NIPS. 2016.
[25] YANG C, WU Q, WEN Q, et al. Towards out-of-distribution sequential event prediction: A causal treatment[J]. Advances in neural information processing systems, 2022, 35: 22656-22670.
[26] ZHOU Z, HUANG Q, YANG K, et al. Maintaining the status quo: Capturing invariant relations for ood spatiotemporal learning[C]// Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023: 3603-3614.



LIU Yuting, born in 1999, postgraduate. Her main research interests include urban flow prediction and spatio-temporal data mining.

GU Jingjing, born in 1983, professor, Ph.D supervisor, is a member of CCF (No. 52397S). Her main research interests include data mining, urban computing and intelligent systems.

(责任编辑: 柯颖)