

基于教师-学生时空半监督网络的 城市事件预测方法

周正阳^{1,4}, 刘浩^{2,4}, 王琨^{3,4}, 王鹏焜^{3,4}, 王旭^{3,4}, 汪炆^{1,2,3,4}

(1. 中国科学技术大学计算机科学与技术学院, 安徽合肥 230031; 2. 中国科学技术大学软件学院, 安徽合肥 230041;
3. 中国科学技术大学大数据学院, 安徽合肥 230031; 4. 中国科学技术大学苏州高等研究院, 江苏苏州 215125)

摘要: 离散时空事件预测是城市计算领域中的重要科学问题之一. 现有工作主要聚焦于使用多样化的时空神经网络对城市动态特征与事件时空关联进行建模, 且已经取得了一定成效, 但仍然存在以下问题: 首先, 城市事件具有诱因多源和时空稀疏性, 而这种时空稀疏性可能同时源于事件本身的稀少性和采集的不完整性, 现有工作尚未能解决短期预测中的稀疏性挑战及零膨胀问题; 其次, 已发生事件倾向于继续向周边区域传播事件风险, 但由于现有工作同质化了动态特征和事件之间的交互关联, 因此其不能捕捉历史事件对未来事件风险带来的交互影响. 鉴于此, 为协同地利用事件标记信息和时空特征, 本文提出基于教师-学生时空半监督学习框架以预测短期离散事件的时空分布. 在教师网络中, 为应对事件标记的稀疏性, 本文在时空学习中引入半监督机制, 提出基于自编码器的特征重建和时空方差异常描述引导的动态特征表示学习; 在学生网络中, 本文设计了特征-事件解耦的双管道学习机制, 并提出时空衰减图卷积网络与长短期记忆网络来模拟事件在时空范围内发生的风险传播. 此外, 本文发展了时空多粒度预测机制, 通过易学的粗粒度预测任务指导细粒度的高质量预测, 最终实现粗-细粒度协同提名的离散时空事件预测. 实验基于纽约和苏州工业园区数据集开展, 本文模型能够在事件击中准确率上分别超越最好的基线模型 5.46% 和 10.65%, 充分验证了提出方法的有效性.

关键词: 事件预测; 时空多粒度预测; 图神经网络; 时空半监督学习; 教师-学生网络

基金项目: 国家自然科学基金(No.62072427); 中国科学院稳定支持基础研究领域青年团队计划(No.YSBR-005)

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2023)12-3557-15

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20211579

A Teacher-Student Spatiotemporal Semi-Supervised Method for Urban Event Forecasting

ZHOU Zheng-yang^{1,4}, LIU Hao^{2,4}, WANG Kun^{3,4}, WANG Peng-kun^{3,4}, WANG Xu^{3,4}, WANG Yang^{1,2,3,4}

(1. School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui 230031, China;

2. School of Software Engineering, University of Science and Technology of China, Hefei, Anhui 230041, China;

3. School of Data Science, University of Science and Technology of China, Hefei, Anhui 230031, China;

4. Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou, Jiangsu 215125, China)

Abstract: Discrete spatiotemporal event forecasting is one of critical scientific problems in the field of urban computing. Existing works mostly focus on leveraging various spatiotemporal neural networks to model spatiotemporal correlations among dynamic urban features and events, and have achieved promising results. However, it still remains the following problems. First, urban events are naturally induced by multiple causes and distributed spatiotemporally sparsely, while such spatiotemporal sparsity can be induced by the inherent infrequent occurrence and its collection incompleteness. Given that, existing works cannot well address such sparsity challenge and zero-inflated issue in short-term forecasting. Second, the occurred events have the potential to raise future risks on neighboring regions. Unfortunately, off-the-shelf literatures tend to homogenize the correlations of feature-event and event-event, and fail to capture the wane-and-wax influences of historical event sequences on future events. Therefore, to cooperatively exploit the event labels and spatiotemporal features,

this paper proposes a teacher-student spatiotemporal semi-supervised learning framework, addressing the challenge of short-term spatiotemporal event forecasting. In the teacher network, to tackle the sparsity challenge of event labels, this paper introduces the semi-supervised scheme into spatiotemporal learning where it designs an AutoEncoder-based feature reconstruction learning and spatiotemporal variance-based anomaly descriptor to facilitate feature representations. In the student network, this work designs a feature-event disentangled dual pipeline and proposes the spatiotemporal attenuation graph convolution network (GCN) and long-short term memory network (LSTM) to imitate the natural risk propagation along spatiotemporal domains. In addition, this paper also develops the spatiotemporal multi-granularity risk prediction task, which emphasizes the easy-to-learn coarse-grained prediction to guide the high-quality fine-grained forecasting, and finally realizes the high-risk discrete region nomination with coarse-to-fine learning mechanism. Experiments on NYC and SIP datasets illustrate that the proposed event forecasting framework outperforms the best baselines by respectively 5.46% and 10.65%, verifying the effectiveness of our work.

Key words: event forecasting; spatiotemporal multi-granularity forecasting; graph neural network; spatiotemporal semi-supervised learning; teacher-student network

Foundation Item(s): National Natural Science Foundation of China (No.62072427, No.12227901); The Project of Stable Support for Youth Team in Basic Research Field, CAS (No.YSBR-005)

1 介绍

城市时空事件预测是城市计算领域的一大挑战,时空事件预测的具体应用包括犯罪预测^[1,2]、传染病爆发预测^[3]、极端天气预测^[4]、交通事件预测^[5,6]等,时空事件预测能够为城市治理如交警警力预配置^[5-8]、道路通行控制等管理行为提供量化指导,同时也能够为个体出行时机选择与路径规划提供合理建议,是数据赋能城市治理的重要方式^[9-12]。

现有的时空预测工作大多针对连续型数据,如基于长短期记忆网络(Long Short Term Memory, LSTM)实现天气^[13,14]、交通流量^[15-22]、空气质量^[23-25]、人类活动与社会经济^[26-31]等预测,而很少有工作针对离散型时空事件进行预测。对于仅有的少量离散型时空数据预测任务,我们将其以预测时间粒度为依据分为长期与短期时空预测两类。长期时空预测是以日为间隔对未来一周进行逐日时空事件分布预测,具体地, Yuan 等人提出了 Hetero-ConvLSTM 网络,该网络通过将离散事件和连续型变量全部同质地嵌入到网格化数据中,并基于 ConvLSTM 提取特征与未来事件的时空依赖关系,最终以集成学习方式得到每个小区域的交通风险模块^[26];类似工作还包括文献^[32],该工作针对逐日道路交通事故风险事件实现了预测,并已集成于 2017 年部署的美国田纳西州交通管理软件;Chang 等基于传统机器学习中的决策树分析了高速公路的事故风险强度^[33];而 Huang 等利用深度学习技术和区域嵌入的方式实现了细粒度长期风险预测^[34]。虽然这些长期序列预测工作已在离散型时空数据预测方面取得了一系列初步成果,但总结来说,这些以日为预测间隔的工作既无法满足市民实时出行的规划需求,也难以应用到实时路网状态风险预测中,因此无法被用来实现短期时空事件预测。近期,面向城市安全的短期时空事件预测

问题越来越被重视。现有小时级别的短期预测模型主要采用 LSTM^[34]、自编码器^[35]或时空注意力机制^[36,37]等方法来实现。然而时空事件的特点在于其时空稀疏性,且这一问题在短周期预测中会愈发突出。因此,基于传统方法的短期预测模型易陷入零膨胀问题,即由于真实值中存在大量零值,损失值最小化的优化目标会迫使学习得到的权重倾向于优化为零,导致预测输出全部为零、无法区分潜在的高风险区域^[38]。此外,城市事件的发生通常源自多种因素,即路网时空特征和历史事件本身均能够对未来事件产生影响,因此,现有的方法将其同质考虑会带来状态特征-事件影响相互交错,使得模型性能陷入次优。综上,考虑到时空事件的稀疏性以及事件间的潜在交互影响,现有时空预测方法仍不能真正解决离散事件短期预测问题。

具体到上述两个重要特性,稀疏性和事件间交互影响,我们对其进一步分析与归纳。对于稀疏性,时空事件预测任务中的稀疏性问题通常是由两类原因造成的:一类是事件或者信息本身密集存在但没有被完整感知,如未安装摄像头或地感线圈的路口将缺失流量信息,我们将这一类原因造成的稀疏称为伪稀疏;另一类是事件或者信息本质上很少发生或存在,如交通事件、犯罪事件这一类偶发事件记录,我们将这一类原因造成的稀疏称为本质稀疏。然而,这两种类型的稀疏并不是各自独立存在的,本质稀疏的信息会同时因为采集不完整,即伪稀疏的原因加剧其稀疏性,这为时空事件预测带来了更大的挑战,使得事件预测问题成为半监督学习范式。对于事件交互影响问题,通常学术界认为城市事件往往是由多因素引发的,其中一个重要且不可忽视的因素是事件本身的时空级联影响^[39]。例如拥堵事件会导致部分车辆超车和频繁变换车道,从而

事故风险以拥堵事件发生地为中心向车辆流动反方向扩散、提升,同时部分车流改变通行路径绕行周边区域会导致周边交通模式改变,提升其他道路事故风险,最终陷入事故-拥堵-事故的恶性循环。另一方面,现有研究表明,在时序上,事件级联影响的传递不仅存在延迟,而且会随着时间的推移逐渐减弱^[40-42],因此单个事件发生后对周边区域会产生“时序上先增强再衰减的风险传递”过程^[3],城市事件也不例外。基于上述分析,我们认为,城市事件具有时空稀疏、离散性强、起因多源且交互影响的特点。因而离散事件预测具有如下挑战:(1)事件序列在时空维度上均存在稀疏性,叠加伪稀疏会加剧这种数据不平衡性,导致预测陷入零膨胀问题^[38],如何从无标记或丢失标记的样本中挖掘潜在事件成为挑战之一;(2)拥堵或事故等异常事件会不可避免地在空间邻近区域和邻接时间间隔内产生影响并叠加风险。因此,如何从时空视角建模事件“全生命周期”(到来、叠加和结束)对周边区域风险带来的时空波动,成为挑战之二。

为解决离散事件短期预测问题所面临的两大挑战,本文以稀疏城市交通事件预测为应用背景来对稀疏离散事件短期预测问题展开深入研究:将城市网格化为若干小区域,进而基于历史交通事故和路网特征、天气等动静语义信息来对每个小区域未来短周期的交通事故风险强度进行预测。具体而言,在城市事件预测中,事件的稀疏性往往源于两种情况:(1)事件发生数量本质较少;(2)部分事件未被感知采集。因此,我们将城市稀疏事件预测建模为半监督学习任务,即部分事件已标记,部分事件因未采集而未标记^[43]。基于此,本文提出了基于教师-学生时空半监督预测模型来实现城市事件风险预测。教师网络表征能力强、学习信息广泛,主要承担挖掘无标记和弱标记信息潜在规律的角色;而学生网络则主要承担与问题更直接相关的学习任务,即监督学习。总体而言,教师网络通过学到更广泛的规律指导“学生”进行细粒度学习,以协同提升预测性能^[43]。为克服零膨胀问题,本文利用教师网络强大的无监督学习机制与表征能力,挖掘稀疏时空数据样本的潜在规律;为充分利用已有数据的统计信息引导教师网络实现异常感知,本文设计了基于时空方差的异常标注机制;为解耦特征-事件的复杂交互关联,本文在学生网络中提出了特征-事件解耦的双管道学习机制和时空多粒度风险预测方法,两个学习管道分别建模特征-事件、事件-事件的映射关联,并基于门控机制实现图神经网络的交互建模,同时发展了时空多粒度预测以缓解零膨胀问题。为解决不同任务之间的优化矛盾,本文提出基于损失值量级的多任务权重自适应策略,以加速模型训练。总之,本文基于教师网络和学生

网络交替优化,一方面缓解了标记稀疏带来的学习不充分问题,另一方面充分利用已有动态特征来实现高质量状态表示,为稀疏离散事件短期预测提供保障。本文贡献有以下几点。

首先,本文首次系统分析了城市事件预测任务的特点及其形成原因。本文指出城市事件的特点及其预测挑战性在于事件的离散性和稀疏性,进一步分析了事件产生原因的多源性,指出多源性主要在于事件和已知多源信息间的交互性及事件影响的传递性。

第二,在技术上,本文首次将教师-学生半监督学习机制引入时空预测,以无监督-有监督交替优化的方式实现了稀疏与密集信息的联合利用。面向事件离散性和稀疏性,我们在教师网络中提出了时空方异常描述为未标记特征指派“事件”伪标签,以无监督学习方式提升特征表达能力和事件挖掘能力;同时在学生网络中设计了时空多粒度预测损失,通过易学的粗粒度风险分布引导复杂细粒度分布学习,缓解空间稀疏性和零膨胀问题。面向异常事件起因的多源性,本文提出了特征-事件解耦的时空交互图网络,解耦了路网车流状态特征与事件、事件与事件之间的交互影响,提出空间距离衰减的图卷积网络(Spatial distance-Decay Graph Convolution Network, SD-GCN)和时间衰减LSTM,即Gamma-LSTM来分别捕获事件在空域和时段上的传播模式与风险趋势。

第三,在实现与实验上,本文提出了基于损失值量级的多任务权重自适应策略,将繁琐的调参过程转化为模型自适应调节任务间权重。我们在纽约和苏州园区两个数据集上验证了模型性能的优越性。实验表明,所提出的模型在事件击中准确率上分别优于最好的基线模型5.46%和10.65%,后续的消融实验、案例分析进一步讨论了模型预测行为的内在机制及其解释性。

2 问题定义

定义1 城市区域和层次型城市区域划分:将城市区域以层次型划分为粗粒度中等矩形区域与细粒度小区域。设全城可划分为 n 个细粒度小区域,并将细粒度小区域的集合定义为 $V=\{v_1, v_2, \dots, v_i, \dots, v_n\}$,按照地理邻接的方式将小区域 v_i 聚合,形成中等矩形区域。设全城可组成 Q 个中等矩形的粗粒度区域,记作 $\text{Reg}=\{\text{rg}_1, \text{rg}_2, \dots, \text{rg}_i, \dots, \text{rg}_Q\}$ 。

定义2 路网动态交通特征与静态路网特征:任意一细粒度区域 v_i ,均包含本区域静态路网特征和动态交通相关特征。区域 v_i 在时间段 t 内的动态交通特征 $f_{v_i}(t)$ 由流量、速度和风险值组成,即 $f_{v_i}(t)=$

$\{TV_{v_i}(t), a_{v_i}(t), r_{v_i}(t)\}$, 其中风险值 $r_{v_i}(t)$ 是该时间间隔内区域 v_i 所发生事件类型与对应类型严重级别的乘积和, 按照惯例^[3, 25, 38], 事件类型分为财产损失、人员受伤和人员死亡三种, 对应于风险等级为 1, 2, 3. 整个路网的静态特征可以定义为一个路网结构分布构成的向量 $S = \{s_{v_1}, s_{v_2}, \dots, s_{v_n}\}$, 其中区域的静态特征 s_{v_i} 由该区域中的道路类型、车道数、道路等级、POI 分布等构成.

定义 3 城市图: 构建无向的城市图 $G(V, E)$. 其中, 城市细粒度小区域集合 $\{v_i | i = 1, 2, \dots, n\}$ 对应于城市图中的节点集合 V . 城市中第 i 个区域与第 j 个区域间的关系用边 $e_{ij} \in E$ 刻画, 其中 e_{ij} 是邻接矩阵 A_e 的元素, 其计算方式如下

$$e_{ij} = \begin{cases} 1, & \text{如果 } v_i \text{ 与 } v_j \text{ 中有路段相连} \\ \sin_{\cos}(s_{v_i}, s_{v_j}), & v_i \text{ 与 } v_j \text{ 中无路段相连} \end{cases} \quad (1)$$

在式(1)中, 若 G 中的区域 v_i 和 v_j 具有邻接路段, 则 $e_{ij} = 1$. 考虑到静态路网结构接近的区域拥有相似的交通模式和交通事故模式, 我们基于区域静态路网特征分布计算区域间相似性, 并将该相似性作为区域间的边权值. 由此, 得到城市图的邻接矩阵 A_e .

问题 1 时空多粒度离散型事件序列预测. 给定整个城市的静态路网结构特征 S 和历史的动态交通信息 $F(t) (t = 1, 2, \dots, T)$, 本文任务是预测未来 r 步上全城事故总数 $O_S(\Delta t) \in \mathbb{R}^1$ 、粗粒度区域离散型事故风险向量 $O_S(\Delta t) \in \mathbb{R}^Q$ 、细粒度区域的离散型事故风险向量 $O_F(\Delta t) \in \mathbb{R}^n$, 及最可能发生事故风险的 M 个细粒度子区域集合 $V_M(\Delta t)$, 其中 $\Delta t = T + 1, T + 2, \dots, T + r$.

3 基于教师-学生网络的时空半监督离散事件预测

3.1 方法概要

鉴于稀疏离散事件是伪稀疏与本质稀疏共存的一种标记稀疏问题, 本文将该问题转化为半监督时空学习任务, 提出基于教师-学生网络的时空半监督离散事件预测模型. 为应对上述“时空事件的稀疏性”与“特征-事件之间的交互级联影响”这两大挑战, 本文分别设计了时空半监督学习机制与特征-事件解耦的时空交互图网络框架. 首先, 本文通过引入半监督学习机制以充分利用未标记信息的潜在规律与内在价值, 在教师网络中实现动态特征重构和基于时空方差的异常表征学习, 以无监督方式提升时空特征的表达能力, 缓解时空稀疏性问题; 其次, 在学生网络中构造特征-事件及事件-事件解耦映射函数, 挖掘离散事件序列形成的映射关系, 同时设计多粒度时空学习算法, 以进一步缓解零膨胀问题并提升预测准确性, 最终以教师-学生网络协

同优化的形式实现权值迁移与知识共享. 本文提出的基于教师-学生网络的时空半监督离散事件预测模型如下图 1 所示.

3.2 基于时空半监督学习的教师网络

受文献[19]启发, 在小样本监督学习中引入无监督学习可以增强特征表达能力. 因此, 教师网络主要以无监督或弱监督形式实现, 协同学生网络进行精细化学习. 本文模型中的教师网络主要由两个部分构成, 分别为动态特征重构和基于时空方差的异常描述伪标记指派. 前一模块将可大量获得的路网车流状态特征(包括速度、流量)进行重构来挖掘时空序列表征, 后一模块则基于无监督式时空方差异常感知增强时空网络的表达能力.

为模拟城市交通数据的时空演变过程, 本文基于定义中构建的城市图网络, 分别改进了 GCN 和 LSTM 算法来实现空间关联挖掘与时间趋势聚合.

首先, 将速度和车流量压缩为一个两通道的张量 $H = [TV; a] \in \mathbb{R}^{T \times n \times 2}$, 通过堆叠图卷积神经网络来逐层传递区域间动态特征的潜在关联. 按照时空计算领域的惯例, 抽取周期性、邻近性和趋势性各 3 个时间间隔的信息进行聚合, 每一个片段均通过两层堆叠的图卷积网络. 其中一个片段的图卷积计算如下:

$$H_{\text{hid}} = \widetilde{A}_e (\widetilde{A}_e H^0 W_{f_0}) W_{f_1} \quad (2)$$

H_{hid} 为空间聚合后的全路网状态向量, 其中 $\widetilde{A}_e = \widetilde{D}_e^{-1/2} A_e \widetilde{D}_e^{-1/2}$, \widetilde{D}_e 为 A_e 对应度矩阵的归一化形式. H^0 为某一片段的二通道张量, W_{f_0}, W_{f_1} 为可学习的 GCN 权重向量. 每一个片段经 GCN 网络聚合后, 均经过 LSTM 进行时序关系聚合, 路网状态的最终表示为

$$H_S = \text{LSTM}(H_{\text{hid}}^0, \dots, H_{\text{hid}}^8; W_{\text{LSTM}}) \quad (3)$$

进而, 本文进一步对已获得特征表示 H_S 进行基于预测重构, 其中 LSTM 可视作编码器、Reconstruct 函数可视作解码器, 得到重构后的特征图 H_F :

$$H_F = \text{Re construct}(H_S) \quad (4)$$

将 H_F 与原始流量、速度序列 H 进行比较, 无监督学习任务的损失函数可写作:

$$\text{Loss}_{\text{self}} = \min \|H_F - H\|^2 \quad (5)$$

另一方面, 由于伪稀疏与本质稀疏均大量存在于事件标注中, 因此需要以特征数据驱动辅助学习, 引导教师网络学习到异常感知的特征表示. 考虑到事件的发生源于状态波动, 而方差是度量数据波动程度的统计量, 因此本文提出基于时空方差的异常描述来对时空特征进行异常标注. 考虑到不同区域的流量和速度具有不同的波动区间和波动幅度, 本文设计了一种基于中位数的区域自适应阈值来判别潜在异常.

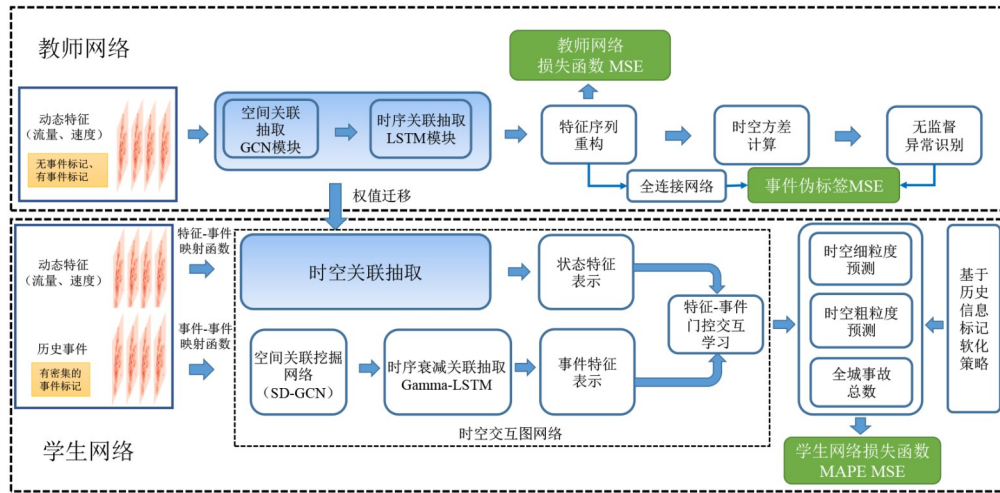


图1 基于教师-学生网络的时空半监督离散事件预测模型

时空数据具有时序上的邻近性、周期性及空间上的邻近性,鉴于此,本文提出将时空方差定义为相似场景下数据波动的度量,波动越大,越可能发生潜在的事故,如车流量陡然增加预示着拥挤事件到来,而流量突然减小可能意味着已发生拥挤或事故,或是周边路网状态的模式发生改变.具体而言,时空方差的实现是基于时序邻近性、周期性和空间邻近性三个不同视角,选取相似场景动态特征数据构成集合,计算其标准方差.其计算机制为:将时序数据以时间步为最小单元组织成以多个连续时间步构成的片段,在第*i*个区域*v_i*的第*m*个时间片段上,对于空间邻近性,首先检索区域*v_i*对应的邻居集合 $\mathcal{N}(v_i)$ 得到空间邻近性对应特征的波动:

$$\text{var}_{\text{Spatial}}^{i,m} = \text{stdv}(\{H(k,m)|v_k \in \mathcal{N}(v_i)\}) \quad (6)$$

对于周期性,从数据中检索时间片段*m*对应的周期性数据集合:

$$\text{var}_{\text{Period}}^{i,m} = \text{stdv}(\{H(i,b)|b \in \text{Period}(m)\}) \quad (7)$$

同理,对于时间邻近性,检索其对应的时间邻近数据集合:

$$\text{var}_{\text{Close}}^{i,m} = \text{stdv}(\{H(i,b)|b \in \text{Close}(m)\}) \quad (8)$$

最后将其融合得到:

$$\text{var}_{\text{ST}}^{i,m} = \frac{1}{3}(\text{var}_{\text{Spatial}}^{i,m} + \text{var}_{\text{Period}}^{i,m} + \text{var}_{\text{Close}}^{i,m}) \quad (9)$$

其中 $\text{Period}(m)$ 和 $\text{Close}(m)$ 均为检索得到的对应时间帧索引集合.例如,在 $\text{Period}(m)$ 中,该时序数据的周期为*T*,那么周期性时间片段为 $b \in \{m - kT|k = 1, 2, 3\}$.

为判断该区域是否发生潜在异常,本文设计了基于中位数的局部自适应阈值来自主标记可能发生城市事件的区域.对于每一区域,计算历史记录中事件发生时邻近时空方差的数值,将方差中位数作为该区

域在测试阶段的自适应阈值,若该区域的时空方差大于该区域的方差中位数,则将其事件发生的可能性伪标记记为1,即认为该区域为活跃区域, $\text{He}_{\text{pse}}(i) = 1$.基于时空方差的异常描述与事件伪标记标注过程如图2所示.

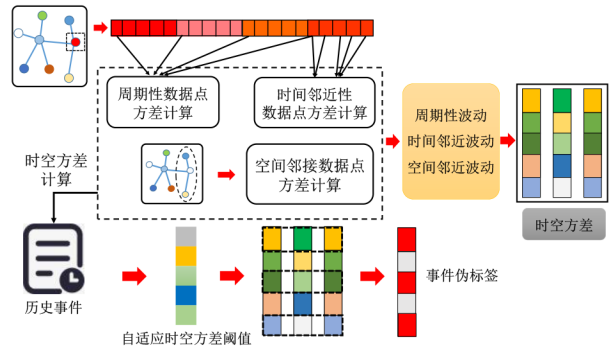


图2 基于时空方差的异常描述与事件伪标签标注

最后,将伪标记传递至教师网络,通过全连接网络引导教师网络感知潜在异常,记为,

$$\widehat{\text{He}}_{\text{pse}} = H_F \cdot w_{\text{pse}} \quad (10)$$

其中, $\widehat{\text{He}}_{\text{pse}}$ 为学得伪标记结果, w_{pse} 是可学习权值.最后得到教师网络联合的优化目标 Loss_{Tea} ,其包括两个方面,一是连续型特征重构误差,二是伪标记学习:

$$\text{Loss}_{\text{Tea}} = \text{Loss}_{\text{self}} + \text{Loss}_{\text{pse}} = \|H_F - H\|^2 + \|\widehat{\text{He}}_{\text{pse}} - \text{He}_{\text{pse}}\|^2 \quad (11)$$

至此,时空半监督教师网络已成功引入无监督和基于伪标记的弱监督学习机制.其具体实现是,将教师网络所学到的GCN网络与LSTM权重 $W_{f_0}, W_{f_1}, W_{\text{LSTM}}$ 进行保存和迁移,作为后续事件预测任务中动态特征提取的权值.

3.3 特征-事件解耦学习的学生网络

学生网络承担目标相关的任务,在本文中即为事件风险预测. 具体而言,鉴于城市交通事件的发生与城市路段的实时状态和历史事件均密切相关,而现有的事件预测任务往往忽略了事件发生后在时空域上存在的持续且动态的影响. 因此,本文提出了基于双管道学习的时空交互图网络(如图1所示),显式解耦了离散事件与连续型特征间的关联,即解耦路网动态特征-事件与事件-事件的纠缠关联,并通过改进GCN与LSTM建模了事件本身产生的交互和级联影响.

基于上述分析,本文提出的学生网络预测模型由四部分组成,教师网络迁移的动态特征-事件映射、时空影响传播感知的事件-事件映射、特征-事件交互学习模块和时空多粒度风险输出模块. 考虑到在教师网络中已经实现了流量、速度等特征的时空图建模,学生网络将直接迁移教师网络中的 $W_{f_0}, W_{f_1}, W_{LSTM}$ 得到的动态特征表征记为 H_S ,以实现教师-学生网络的知识共享. 而特征-事件交互学习模块将解耦学习的特征表征和事件影响表征进行融合,时空多粒度风险输出模块将易学的粗粒度预测任务指导细粒度的高质量预测,进一步缓解稀疏性带来的零膨胀问题. 接下来,我们将逐个讨论时空影响传播感知的事件-事件映射、特征-事件交互学习与时空多粒度风险输出模块.

3.3.1 基于SD-GCN和Gamma-LSTM的事件时空影响传播模型

在城市中,当某一事件发生后,该事件所产生的风

险将沿着相应的时空域进行传播,进而诱发下一事件,这种关联即为事件的交互与级联影响.

具体地,在空间方面,如图3(a)所示,区域A(位于十字路口)、区域B(位于三叉路口)分别发生一起交通事故. 区域A呈现同心圆式,车辆积压与拥堵传播,区域B呈现同方向风险扩散与事故叠加,表明事件以发生点为中心的向邻近路段扩散风险影响,且这种影响将随距离增加而逐渐减弱. 本节将全城已发生事件的风险记为 $H_r^0 \in \mathbb{R}^{T \times n \times 1}$,并基于静态路网结构构建一个具有空间距离衰减的事故风险传播图,区域 v_i, v_j 的边 e_{ij}^r 构成了基于空间距离衰减的邻接矩阵 \tilde{A}_r ,按如下式(12)计算:

$$e_{ij}^r = \exp(-\text{dist}(v_i, v_j)) \quad (12)$$

因此,基于空间距离衰减的风险传播如式(13):

$$H_{\text{hid}}^r = \tilde{A}_r (\tilde{A}_r H_r^0 W_{r_0}) W_{r_1} \quad (13)$$

其中, H_r^0 为事故风险向量.

在时序方面,如图3(b)所示,区域A单一事件随时间增加车辆逐渐积压,其对于道路风险的影响先增加,而后随事件处理结束而逐渐衰减. 区域B在时间步4产生了事件叠加,最终展示了事件叠加后对该B区域的风险影响. 由于事件在发生后存在短期车流积累过程,并不会立刻导致拥堵. 因此事件发生后,其对当前区域的影响呈现先增大再缓慢降低的过程. 而在近邻区域中,事故共现可能产生风险叠加,导致陷入“事故-拥堵-事故”循环,图3(b)中的区域B展示了这一过程.

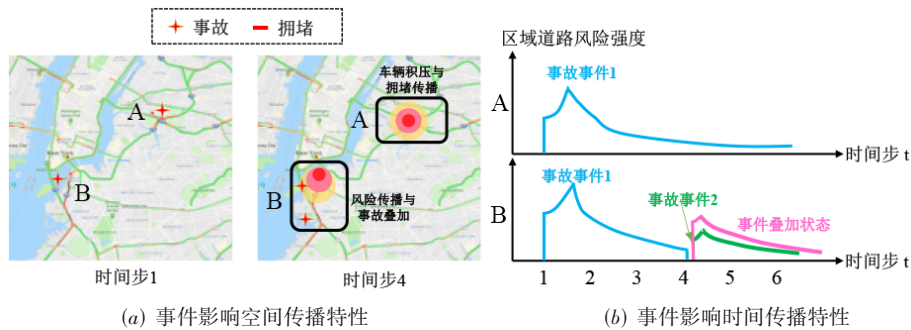


图3 异常事件在时空域上的风险传播.

本文研究发现,如图4所示,Gamma分布在 $\alpha > 1$ 情形下恰好描述了事件发生后周边区域风险随时间消长的过程,

$$f_{\text{gamma}}(x, \beta, \alpha) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad (x > 0) \quad (14)$$

其中,其均值和方差由参数对 (β, α) 刻画,均值 $\mu = \frac{\alpha}{\beta}$, 方差 $\sigma^2 = \frac{\alpha}{\beta^2}$. 经实验论证,本文选取 $\alpha = 2.0 \sim 3.0, \beta =$

0.4~0.6来刻画事件-事件的概率分布.

因此,在时序关系上,将Gamma分布引入时间序列预测模型LSTM,以建模历史事件对当前及未来风险状态的消涨趋势. 将Gamma分布视作一个门控单元,并记为 $g_{GM}(e_{v_i}, t_i) \leftarrow f_{\text{gamma}}(x)$, 其中 t 为原Gamma分布中的 x . 由于事件的连续性影响,Gamma门控仅作用于最后3个连续事故风险序列. 在本文中,我们提出为每一个出现在张量 R 中的事件 e_{v_i} 分配一个时间步计数器 tp_i ,每经过一步其对应的时间戳将加1. 值得注意的是,

在解码部分,本文通过引入时间步引导的Gamma-LSTM (LSTM_{Gamma})来模拟事件影响的涨逝.具体地,将时序建模部分记为

$$H_R = g_{GM}(ev_i, tp_i, \alpha, \beta) \odot LSTM_{Gamma}(H_{hid}^{r,0}, \dots, H_{hid}^{r,8}; W_{Ga-LSTM}) \quad (15)$$

其中 \odot 是按位相乘, α, β 是可调节的超参数, $W_{Ga-LSTM}$ 是可学习参数, $H_{hid}^{r,0} \sim H_{hid}^{r,8}$ 表示共9个时间步.

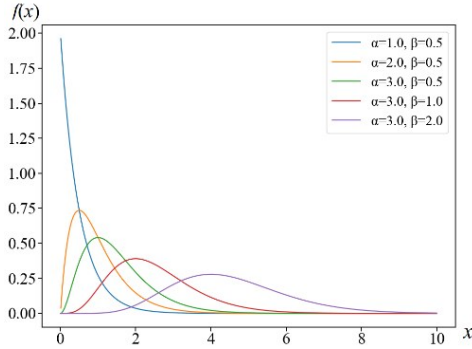


图4 Gamma分布

3.3.2 特征-事件门控交互学习机制

至此,我们获得了路网动态特征和事件风险传播感知的时空表示.为融合路网状态和事件风险,本文提出一个门控单元来实现状态融合.具体地,将其将9个时间步的编码和对应时间步的天气等语境信息作为门控单元的输入,记为 $Conxt_i$,并选用Sigmoid激活函数实现门控单元.对于多步预测,将第*i*步的时间戳和天气分别记为 tp_i 和 $weat_i$,而上下文语境向量 $Conxt_i$ 可记为时间戳和天气的级联向量.进而,本文有:

$$Conxt_i = \text{concat}([tp_i; weat_i]) \quad (16)$$

$$\gamma_i = \text{Sigmoid}(Conxt_i) \quad (17)$$

其中, γ_i 为门控单元输出的融合权重.而后基于 $Conxt_i$ 可得时间步和天气信息感知的特征-事件风险聚合表示,

$$HR_i = \gamma_i H_s^i + (1 - \gamma_i) H_x^i \quad (18)$$

进而,我们实现了语境因素与状态的实时交互与融合.

3.3.3 事件风险标签软化技术与上下文引导的多粒度事件风险预测

面对短期时空事件预测存在的时空稀疏性与零膨胀挑战,我们有如下两个观测:(1)在对全城细粒度小区域风险值排序保持不变的情况下,对风险值逐一变换得到一系列具有区分度的值,能够使网络易于训练;(2)在时空预测问题中降低预测的时空粒度,可提升时间步和网格中的非零值覆盖率,进而提升总体预测性能.因此,我们基于文献[3]的相关技术,进一步发展了标签软化技术和多粒度风险预测机制,以共同克服零

膨胀问题、使数据更易于神经网络学习.

(1)事件风险标签软化技术

考虑到离散事件标记的信息量较小,我们设计了基于历史信息的标记软化机制,基于间隔最大化的思想对标签进行变换,将事件风险值变换为一个正负分明的实数值,而非(0,1)区间的标准概率数值.具体而言,当该时间段内某区域未发生事件,则用负数表示该区域暂无风险,当该时间段内某区域确有事件发生,则用发生事件等级之和的正数表示,事件数量越多、越严重,其数值越大.具体操作是将训练集中各个区域的风险总数求和,再归一化至(0,1)之间,进而利用对数函数在(0,1)区间内具有的负数变换、单调递增、增强数值区分度等特点,实现数据对数化.具体地,设 N_{day} 表示采样样本中的天数,对区域 v_i 的历史风险值进行(0,1)区间归一化可获得各区域的强度指示 ε_{v_i} :

$$\varepsilon_{v_i} = \frac{1}{N_{day}} \sum_{j=1}^{N_{day}} \frac{r_{v_i}(j)}{\sum_{k=1}^m r_{v_k}(j)} \quad (19)$$

进而对数化得到对应的负数 π_{v_i} ,

$$\pi_{v_i} = b_1 \log_2 \varepsilon_{v_i} + b_2 \quad (20)$$

其中, b_1, b_2 为两个可调节的参数,用于约束正负风险数值在对等范围内.统计发现,风险值一般为0~5区间,那么应使 π_{v_i} 的取值范围在-5~0区间,我们根据定义域 $\varepsilon_{v_i} \in [0.1, 0.9]$ 得到一组符合条件的参数($b_1 = 5, b_2 = 0$),其变换过程如图5所示.

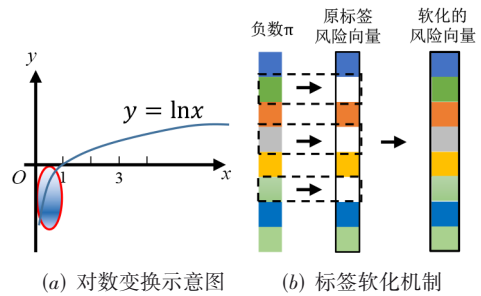


图5 标签软化技术

在训练学生网络时,当标记的风险数值为0时,将该0值替换为对应的 π_{v_i} .值得注意的是,由于特征重构与事件风险预测两个任务可能存在潜在的优化冲突,因此交替训练教师、学生网络,以使得两个网络逐渐收敛达到整体最优解.

(2)上下文引导的多粒度事件风险预测

多粒度事件风险预测同时进行细粒度和粗粒度时空预测,即同时输出如下信息:(1)细粒度多步风险预测 O_F ;(2)降低预测粒度,得到时-空上均为较粗粒度的多步风险预测 O_C ;(3)在每个细粒度时间步中,预测每

个时间步中全城事故总数 O_S 。

具体地,我们将HR视作时空细粒度预测表征,即 $O_F=HR$,对于多步预测中的每一步 i 有 $o_{f_i}=HR_i$ 。同时,由于空间粒度降低,数据稀疏程度减缓,因而我们基于全连接神经网络实现从细粒度空间到粗粒度空间的局部区域风险聚合,实现细粒度-粗粒度信息交互学习。特别地,文献[33]与[35]表明,上下文语境信息与事故风险具有较强相关性,因此,本文将上下文语境信息融入粗粒度信息中进一步引导粗粒度风险信息准确聚合:

$$o_{c_i} = o_{f_i} \cdot w_{f2c} + \text{Conxt}_i \cdot w_{\text{con}2c} \quad (21)$$

其中 $w_{f2c}, w_{\text{con}2c}$ 分别为两组可学习权重,且每一预测时间步均共享这两个权重。我们将这两组表征以按位相加的方式进行融合。

而后,我们进一步对数据进行重组以获得层次性的时空信息 o_{stc} 。具体而言,在时序上将时间步以两步为一组进行聚合,在空间上按照矩形区域数据进行整理,得到粗粒度的时空风险:

$$o_{\text{stc}_i} = \mu o_{c_{2i}} + (1 - \mu) o_{c_{2i+1}} \quad (22)$$

为简便起见,本文假设时空细粒度预测的时间总步数为偶数,且超参数 $\mu=0.5$,关于超参数的设定将在4.6节进行讨论。将粗粒度时空风险进一步聚合以可学习形式获得全城事故总数 $o_{S_i} = o_{c_i} w_{c2s}$,其中, w_{c2s} 是空间粗粒度到全城事故总数的变换函数,学习得到的多粒度输出 o_F, o_S, o_{stc} 则为层次性的时空信息。将这些层次性的多粒度输出分别与重组得到的真实值计算损失函数,分别记为 $\text{Loss}_F, \text{Loss}_C, \text{Loss}_S$,进而得到学生网络的联合优化目标:

$$\text{Loss}_{\text{stu}} = \eta_F \text{Loss}_F + \eta_C \text{Loss}_C + \eta_S \text{Loss}_S \quad (23)$$

然而对上式中的任务权重进行超参数调优是较为困难的。为平衡各项任务、提升总体优化效率,本文提出任务自适应权重匹配机制。我们认为,在优化过程中,损失差异变化越小表明通过学习所能获得的提升空间越小,而差异变化较大则表明其优化尚不充分。因此自适应权重计算考虑了任务损失量级、训练过程损失函数实时衰减速率等因素。首先,将各任务权重初始化为各自初始损失占总损失的比重,以 η_F 为例,

$$\eta_F^0 = \frac{|\text{Loss}_F^0|}{|\text{Loss}_F^0| + |\text{Loss}_C^0| + |\text{Loss}_S^0|} \quad (24)$$

其中 Loss_i^i 表示子任务*在训练过程中第 i 批(batch)的损失值。随着训练批次增加,各子任务权重应与该任务损失变化的差异成正比,变化越大则在该阶段越应当重点优化,因此,第 $i+1$ 批(batch)的任务权重以式(25)更新:

$$\eta_F^{i+1} = \frac{\Delta \text{Loss}_F}{\Delta \text{Loss}_F + \Delta \text{Loss}_C + \Delta \text{Loss}_S} \quad (25)$$

其中, $\Delta \text{Loss}_i = |\text{Loss}_i^{i+1} - \text{Loss}_i^i|$ 表示训练过程中第 i 批与上一批相比损失函数的变化,衡量了对应于任务*学习的快慢。最后,联合式(23)~(25),融合天气、时间等语义信息引导空间粗粒度事件风险学习(式(21)),将时空细粒度与粗粒度协同预测。这三项任务可视为互为正则化项,共同促进预测性能与鲁棒性提升。

与文献[3]相比,本节的贡献在于明确和凝练了基于对数函数进行标签软化的动机和参数选取方法,在多粒度学习中进一步提出了多粒度时空数据构建方法,并提出自适应任务权重调节机制,完善和发展了多粒度事件风险预测,提升其完整性和智能性。

3.3.4 基于分块池化的离散序列生成

融合层次性事件分布,全城事件总数 O_S 、粗粒度事故风险分布 O_C 、细粒度风险分布 O_F ,实现离散高风险事件自适应筛选。对于第 i 个时间步,按照中等矩形区域进行分块划分,基于函数 Select 实现全城可能发生事件的区域提名:

$$V_{m_i} = \text{Select}(o_{S_i}, o_{c_i}, o_{f_i}) \quad (26)$$

具体地,首先将每一时间步中 O_{S_i} 作为全城事件高风险区域的个数,而后按照学得 O_C 在空间上的分布,按比例分配第 j 个中等矩形区域所包含的高风险小区个数 $q_i(j)$,记为

$$q_i(j) = \text{int}\left(O_{S_i} \frac{o_{c_i}(j)}{\sum_{j=1}^Q o_{c_i}(j)}\right) \quad (27)$$

最后,基于细粒度风险表征的 $\text{Top}-q_i(j)$ 区域索引构建事件发生区域的预测集合,以实现函数 Select 。时空多粒度事故风险预测流程总结如图6所示。

4 实验

本文从性能比较、消融研究、案例分析等多个视角对提出的离散时空事件预测方法进行实验评估。

4.1 数据描述

本文在两个真实数据集上开展实验:(1)纽约交通事故事件(NYC Vehicle Crash Collision, NYC):2017年1月1日至2017年5月31日^①;(2)苏州工业园区事件数据(Suzhou Industry Park, SIP):2017年1月1日至2017年3月31日^②。对于NYC数据集,使用每个小区域的出租车出行量作为人们出行活动的指示;对于SIP,使用

①NYC交通事故数据:<https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95>

NYC出租车旅程信息:<https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

NYC速度数据:<http://data.beta.nyc/dataset/nyc-real-time-traffic-speed-data-feed-archived>

②SIP交通数据集:<https://github.com/zzyy0929/STUaNet>

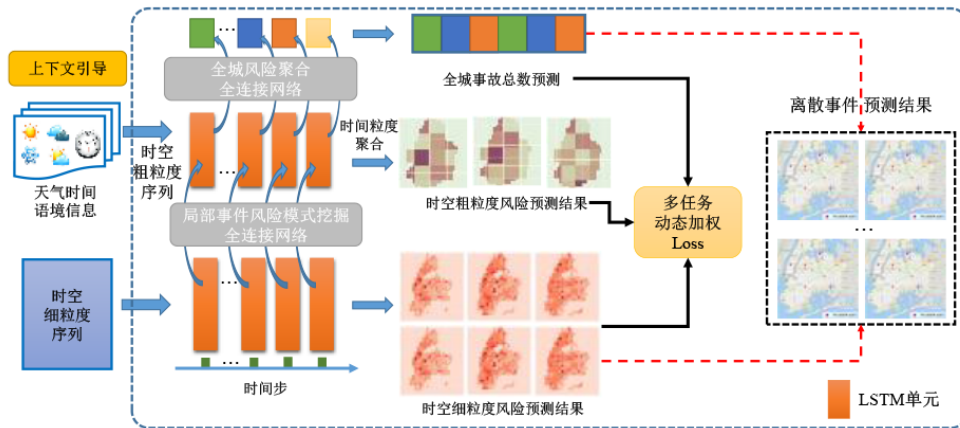


图6 时空多粒度事故风险预测

交通流量、速度和新浪微博收集的交通事故作为实验数据,各数据集统计值如表1所示。

表1 数据集统计量

城市	数据集	时间跨度	区域个数	记录条数
纽约城 NYC	交通事故	2017/01/01- 2017/05/31	354	254k
	出租车 旅程记录			48 496k
	路段速度			125k
	天气			604
苏州工业 园区 SIP	交通事故	2017/01/01- 2017/03/31	108	183
	交通流量			1 399k
	路段速度			311k
	天气			180

4.2 实验设置

4.2.1 实施细节

在数据集划分方面,本文分别选择60%、30%和10%的样本量作为训练、测试和验证集。对于NYC,将NYC研究区域划分为1.5 km×1.5 km的小正方形区域,得到具有路段区域的354个方形子区域。在SIP数据集中,一共有108个路口监控点,即108个区域(节点)。所有的时间间隔均为10分钟。在这两个数据集上,我们将GCN的数量设置为3,每层图卷积过滤器设为320维,对于LSTM模块,在NYC和SIP数据集上隐层单元个数分别设为320和96,均仅使用2个LSTM层。本研究在具有16 GB显存的单个Tesla V100上进行模型训练,基于Python 3.6和Tensorflow 1.14.0构建深度图神经网络。

4.2.2 基线模型

选取如下八个典型的时空预测模型作为本文基线方法。为公平起见,本文对于所有基线模型,均基于历史12个时间步和三个影响因素(即交通量、平均速度和

事故)对未来6个时间步的全城风险进行预测。且除ARIMA外,本文在其他所有基线模型上均进行了标记变换,以缓解深度学习优化过程的零膨胀问题。所有基线模型均根据原始文献的推荐超参数设置进行初始化,并在本文使用的数据集上进行调优,以获得最佳性能表现。我们选取的基线模型如下:

(1) ARIMA是一个基于机器学习的经典时间序列预测算法,其中ARIMA的参数(p, d, q)设置为(1, 2, 6)。

(2) LSTM是一个经典的深度学习时间序列预测模型,在NYC和SIP数据集上,分别使用了256和96个LSTM隐层神经元。

(3) Hetero-ConvLSTM是用于交通事故预测的深度时空学习模型^[25]。NYC的网格被划分为27×27, SIP被划分为15×10,卷积核的大小均为3×3。基于历史最近的12个时间步来预测后6个时间步。

(4) STGCN是一个多步交通预测模型,其包括图卷积和门控时序卷积^[20]。本文堆叠了两个时空卷积模块以捕获空间关联,且每一个过滤器的维度为64。

(5) STG2Seq是一个基于城市图的出租车需求预测模型,其使用一个层次性的图卷积结构来同时捕获时空关联^[21],我们设置了6个具有32维度的GCN过滤器。

(6) STSGCN通过堆叠3个邻接矩阵获得同步时空邻接矩阵,以建模空间同步的动态性与异质性^[22]。本文引入了4个同步图卷积层,每层包括具有64个过滤器的3个图卷积迭代操作。

(7) STDN基于到达流量与流出流量的相互作用实现交通流量预测,并解决了时空预测中的时序偏移问题。我们堆叠了3层CNN且每个卷积核为3×3,包含64个滤波器,LSTM的隐藏维数设为128^[40]。

(8) RiskSeq是目前较为先进的交通事件预测模型,但尚未引入半监督学习,也未考虑事件之间的级联影响^[3]。

4.3 实验评估结果

4.3.1 评价指标

本文选取均方误差(MSE)作为事件风险回归预测的评价指标. 在时空多粒度预测中,MSE-F表示逐区域细粒度预测风险误差,MSE-C表示时空粗粒度预测风险误差.

对于离散事件击中准确率,选取 Acc@M,该指标表示在某一时间间隔中,预测最高风险的M个区域中发生真实事故的区域占真实发生事故区域总数的百分比. 由于在本文设定的时间间隔(10分钟)中, NYC上的平均事故数约为20, SIP上的平均事故数约为6,因此在不具备自适应提名机制的所有基线模型中,采用 Acc@20/Acc@6来进行统一性能评测. 对于本文提出的自适应提名机制,我们在消融实验中采用 Acc@K对结果进行评估,K为学习得到的、时间步相关的全城事故总数.

4.3.2 性能比较

在性能比较方面,我们计算其所有时间步预测误差和准确率的平均值,将其报道于表2中.

表2 基线模型的性能比较

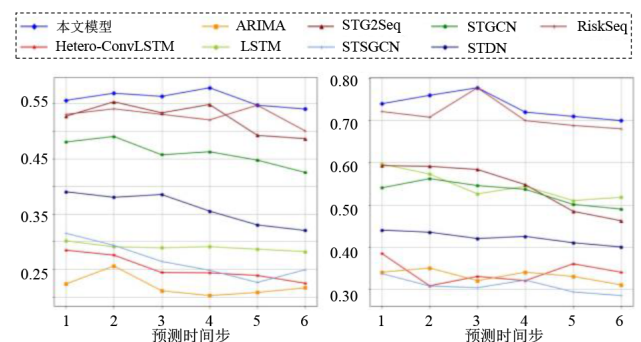
模型	NYC数据集/SIP数据集		
	Acc@20/Acc@6	MSE-时空 细粒度	MSE-时空 粗粒度
ARIMA	20.62%/30.63%	0.019 4/0.018	0.049 5/0.222
LSTM	28.95%/36.55%	0.018 5/0.026	0.048 6/0.027
Hetero-ConvLSTM	28.05%/34.50%	0.016 8/0.049	0.101 2/0.404
STGCN	50.42%/51.37%	0.019 6/0.045	0.049 4/0.288
STG2Seq	52.08%/54.03%	0.014 4/0.364	0.069 3/0.167
STDN	37.48%/42.18%	0.020 5/0.036	0.054 7/0.223
STSGCN	34.75%/40.56%	0.035 4/0.032	0.064 1/0.336
RiskSeq	53.45%/70.27%	0.016 2/0.040	0.044 5/0.270
本文模型	56.37%/77.76%	0.005 5/0.026	0.180 4/0.104

MSE-F和 Acc@M用于衡量细粒度事件预测的性能. 本文提出的模型在所有参与比较的基线方法中达到了最高准确率和较低 MSE 误差. 在 NYC 数据集上,本文的时空半监督网络在 Acc@20 指标上比最佳基线模型提高了 5.46%;在 SIP 数据集中,本文模型达到了 77.76% 的最高准确率,比最佳基线模型高出 10.67%. 该实验结果表明,如果基于纽约市预测风险 Top-20 区域和苏州工业园区的预测风险 Top-6 进行筛选,那么这两座城市有超过 56% 和 75% 的真实事故被准确预测. SIP 数据集上性能相对较高的原因可能在于本身存在的事件较少且多数发生在有限、重复的路口,模式相对稳定. 分析模型及其对应的实验结果,可以发现:深度学习模型可以编码空间和时间相关性而非深度模型

ARIMA 仅考虑了时间依赖性,因此多数深度模型比 ARIMA 表现得更好,且同时实现时空建模的模型优于序列预测模型. 模型 STGCN 和 STG2Seq 分离了长期和短期建模,并采用图卷积和门控机制进行聚合,成功捕获了短期流量变化趋势并获得了比其他基线模型更好的结果. 尽管 Hetero-ConvLSTM 能够实现逐日预测,并同时考虑了空间异质性和层次性,但由于城市区域的不规则性和短期事件的时空稀疏性,该模型仍无法适应多步短期事件预测. 此外,STSGCN 采用扩展邻居矩阵策略带来较高的计算成本. 而本文提出的模型获得较优性能的原因在于:(1)本文从路网动态特征-事件的双视角对事件成因和复杂依赖进行解耦建模;(2)在逐步预测中实现粗、细粒度同时预测并融入更多额外上下文信息;(3)半监督学习提升了时空特征表达能力.

4.3.3 时间序列多步预测的稳定性比较

为评估模型在多步预测中的稳定性,我们通过实验比较了不同模型在预测未来6个时间步上预测的性能,并将结果展示于图7中. 实验显示,模型 STG2Seq、STGCN 的预测性能和本文模型相比较为接近,因为它们都很好地利用了空间和时间上的聚合优势. 特别地,在多步预测中,本文所提出的方法能够在更远期的时间步预测中保持稳定,即使是 NYC 和 SIP 的最后一个时间步也保持超过 53% 和 70% 的事件预测准确率,而 STG2Seq、Hetero-ConvLSTM 等方法在第四个时间步之后性能逐渐下降,并且在最后一步都下降至低于 50%. 以上结果,进一步证实了“事件影响时空传播机制”和“时间步上下文引导的粗粒度学习”在提升多步预测能力上的有效性.



(a) Acc@20 in NYC

(b) Acc@6 in SIP

图7 时间序列多步预测的稳定性比较

4.4 消融实验

为评估各模块对解决问题、性能提升的贡献,本节从特征-事件解耦与稀疏数据预测两个角度进行消融研究,依次消去各子模块来验证其对整体模型的有效贡献度.

4.4.1 特征-事件解耦相关消融研究

为研究特征-事件解耦机制是否更有利于学习任务,本节将事件向量与动态特征向量进行连接操作,以构成多维张量,将多维张量视作整体直接输入传统 GCN 与 LSTM 得到消融模型预测结果.同时,我们分别消去 SD-GCN 和 Gamma-LSTM,观察模型性能变化.我们将这三个消融变体模型分别记为 DiscSeq-NonCas、DiscSeq-SD 和 DiscSeq-Ga,实验结果如表 3 所示.实验发现,去除事件级联影响的时空传播模块后,事故预测性能下降了 12.34%,这表明考虑事件本身存在风险级联影响对未来风险预测具有较大意义.去除“事件影响空间衰减”与“事件影响时序衰减”后,变体模型预测准确率均有所下降,这也验证了“对事件在时空上的衰减传播过程建模”是必要的.

4.4.2 时空稀疏性相关消融研究

针对时空稀疏性,本文提出了基于教师-学生网络的时空半监督学习和时空多粒度预测机制.本节将依次删减教师-学生网络中的时空方差的伪标记标注、时空特征自编码器重构模块、时空粗粒度预测模块,将其消融变体分别记为 DiscSeq-STvar, DiscSeq-self, DiscSeq-Coarse.为验证自适应任务权值分配的有效性,本文还设计了非自适应的等额任务权重变体 DiscSeq-NoAdaw.消融实验结果如表 3 所示.为了从数值实验上进一步验证模型在应对稀疏性方面的有效性,我们在两个数据集上分别进行了稀疏性挑战实验;通过“mask”的方式随机掩去 20%、30%、40% 和 50% 的事件标签,观察不同消融变体在不同的标签稀疏程度下其性能的变化.稀疏性挑战实验结果如表 4 所示.

表 3 消融变体模型性能比较

消融变体模型	NYC 数据集/SIP 数据集		
	MSE	Acc@20/Acc@6	Acc@K
DiscSeq-NonCas	0.012 5/0.038 5	49.52%/72.58%	53.44%/70.23%
DiscSeq-SD	0.013 0/0.036 4	52.41%/68.66%	47.56%/64.32%
DiscSeq-Ga	0.008 4/0.035 7	51.77%/70.45%	50.75%/66.74%
DiscSeq-STvar	0.005 4/0.028 5	49.56%/70.24%	49.17%/68.55%
DiscSeq-self	0.013 7/0.024 3	48.39%/67.42%	48.43%/64.52%
DiscSeq-Coarse	0.009 5/0.025 8	47.67%/63.55%	47.25%/62.36%
DiscSeq-NoAdaw	0.001 0/0.026 5	48.56%/64.30%	49.10%/65.44%
本文模型	0.005 5/0.026 0	56.37%/77.76%	54.05%/73.56%

观察表 3 可以发现,去除粗粒度学习与自编码器重构模块后变体模型性能下降较快,分别在 NYC 数据集中下降了 16.37% 与 14.15%,表明本文所设计的教师-学生网络是合理正确的.去除时空方差引导和统一建模特征事件后,变体模型预测性能也略有下降,表明时空方差的引导和解耦建模能够在一定程度上提升事件预测性能.而在等额任务权重变体中,其学习质量低于自适应权重的学习结果,可表明自适应权重是一种有效提升学习质量的策略.

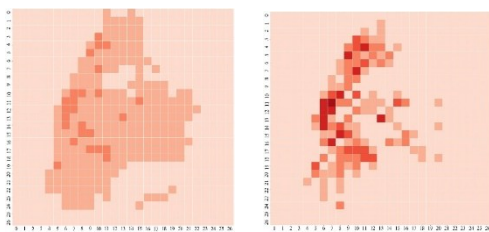
在稀疏性挑战方面,表 4 的实验显示,集成我们各模块的模型具有较强的抗稀疏能力.其中,在 NYC 数据集上去除 30% 和 40% 标记情况下,仍可全面超越两个变体模型且准确率均大于 50%;在 SIP 数据集上,去除 30%、40%、50% 标记时全部优于消融变体模型.但值得注意的是,在去除 50% 标记后,我们的本体模型和变体模型都保持在 45% 左右的准确率,这可进一步说明教师网络的半监督学习机制对模型性能提升的作用是显著的,而此时相对较低的性能可能是由于已标记的信息量较少,导致了全局规律学习不完善.模型对数据的认知完整性欠缺.尽管如此,以上数值结果已能够

说明我们所提出模块的有效性.

表 4 消融变体模型在不同稀疏程度数据集上的性能比较

消融变体模型	NYC 数据集(Acc@20)			
	去除 20%	去除 30%	去除 40%	去除 50%
DiscSeq-Teacher	50.65%	48.42%	46.55%	44.08%
DiscSeq-Coarse	51.52%	50.49%	46.91%	46.25%
本文模型	54.87%	53.53%	50.54%	45.60%
消融变体模型	SIP 数据集(Acc@6)			
	去除 20%	去除 30%	去除 40%	去除 50%
DiscSeq-Teacher	72.45%	68.95%	65.84%	63.22%
DiscSeq-Coarse	73.66%	70.54%	66.72%	64.37%
本文模型	76.76%	73.52%	70.48%	65.16%

为进一步验证多粒度学习机制的有效性,我们以可视化的形式展示了“多粒度学习”的消融结果.具体而言,我们将消融多粒度学习模块前后的风险分布预测结果展示至图 8.结果显示,具有多粒度学习模块的模型输出结果具有更高的区分度,而多粗粒度学习引导的风险分布相对趋同,由此可见,空间多粒度预测可以提升模型风险辨别能力、缓解零膨胀问题.

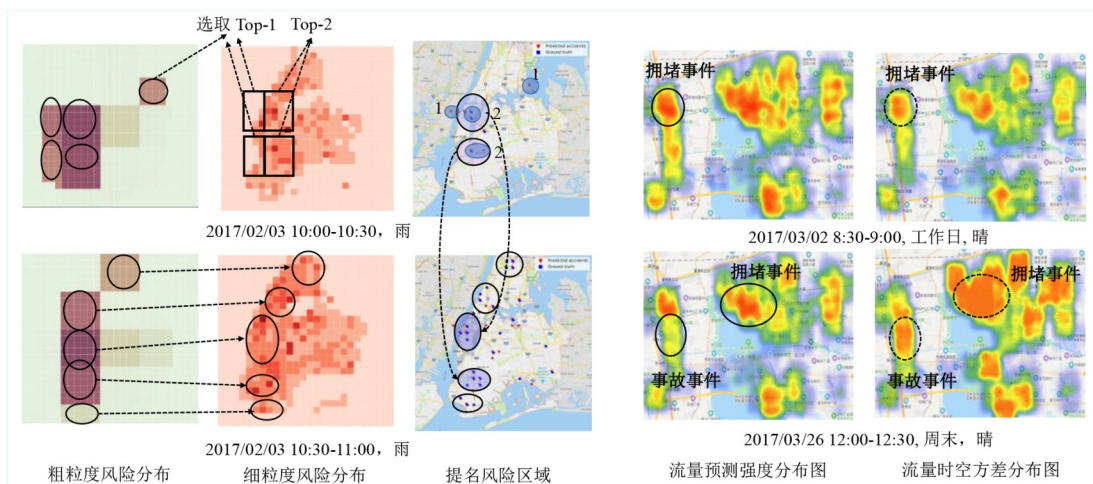


(a) 多粒度预测的模型输出 (b) 具有空间多粒度预测的模型输出
图8 2017年5月12日事故风险分布预测

4.5 案例分析

分别选取NYC和SIP两组测试数据,将模型的预测结果可视化于图9中.图9(a)示意了高风险区域的筛选提名过程,同时可视化了NYC数据集在2017年2月3日10:00~11:00期间多粒度时空预测和事故区域提名结果.在图9(a)的上子图中,可计数得到全城风险区域共7个,其中3个中等矩形区域各需提名1个高风险小区域,而2个中等矩形区域需各提名2个高风险小区域.根据实验案例,有如下观测:(1)预测结果与真实交通事故分布在空间上是相近的,且全城预测总数与粗

细粒度结果能够随天气和时间变化而变化;(2)第三组图中,两条虚线连接的区域在雨天发生了明显的事事故级联现象,即发生某起事故后,事故风险在时空域上扩散,继而在雨天路滑的天气下加剧了事故风险,产生了更大范围的交通事故;(3)全局提名机制可能因交通流量不平衡和空间异质性导致忽略边远地区局部风险地点,本文通过对粗粒度区域分块池化的方式提名局部风险区,可鼓励总体低风险但局部高风险的偏远区域被提名.在图9(b)中,我们分别可视化了SIP数据集上工作日与周末的预测结果,被虚线圈出的是当前流量时空方差较大的区域.结合流量强度可推断其可能发生了交通拥堵或事故,而这也与事实相印证.这些结果表明本文提出的时空方差异常描述是正确有效的.在出行模式方面,我们发现,周末的拥堵(事故)出现在中心商业圈,而工作日拥堵(事故)往往发生在城市通勤的主干道;天气晴好的周末会为中心商业圈带来极大的人流量,而工作日的主干道则不受天气影响,因此不同的上下文往往会带来不同的时空信息聚合方向和截然不同的交通模式,这也验证了上下文引导策略的正确性.



(a) 纽约城(NYC)层次性事故风险分布与可能事故事件提名结果 (b) 苏州工业园区(SIP)流量预测、时空方差与真实事件分布
图9 案例分析

总之,通过上述时空预测的可视化,进一步印证了本文模型设计的正确性与预测结果的解释性.

4.6 超参数敏感性试验与模型效率分析

本文涉及需要调整的超参数分别是时间粗粒度聚合因子 μ 、LSTM层数和隐层单元个数、GCN层数与GCN隐层单元数.在本文研究的两个数据集中,本文根据已有研究^[3],将LSTM层数均设置为2,其余参数通过实验进行优化.我们在图10中展示了NYC数据集上超参数优化过程.最后,我们得到 $\mu=0.5$,LSTM隐层单元数为320,GCN层数为3,隐层单元为256.同理,在SIP数据

集上, $\mu=0.5$ 时也取得最优结果,而最优的LSTM隐层单元数为96,GCN层数为3,隐层单元数为128.

鉴于超参数的设置直接影响模型性能与复杂度两个方面,因此需要在复杂度与性能间取得平衡.假设每个GCN模块中的神经元数量为 n ,SD-GCN中的参数数量为 $O(n^2)$.在模型实现中, n 设置为256,GCN包含3个块.参数总数为 $256 \times 256 \times 3$,而在测试时,平均每轮事故预测需要5.6s,该复杂度和测试时间可满足后台实时多粒度预测要求.

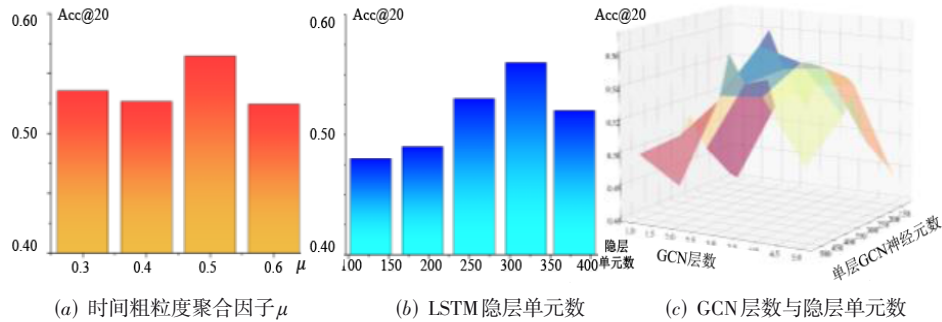


图 10 模型在不同超参数下表现(NYC数据集)

5 结论与讨论

本文提出了一个系统且新颖的、基于时空半监督学习的城市离散事件预测方法。本文核心创新性在于首次将教师-学生网络的半监督学习机制引入时空学习,通过构建无监督学习任务获得强大的教师网络表征能力,并基于稀疏事件的传递性建模事件规律,从而以密集-稀疏信息的协同学习来应对时空稀疏性挑战。具体地,在教师网络中,本文提出了动态时空特征重构与基于时空方差的异常标签指派机制,以引导获得高质量特征表征、提升数据利用率;在学生网络中,本文解耦了动态特征与事件间复杂耦合关系,以挖掘事件本身带来的级联依赖,并通过时空风险衰减来建模事件影响的传播过程,最后以时空多粒度的形式实现了事件风险预测。总之,模型从数据利用、数据解耦、数据训练等视角协同克服时空稀疏挑战,并在两个城市交通事故数据集中进行了验证,其效果超越了所有基线方法。

“教师-学生网络”的设计机制是利用教师模型传授知识至学生模型,形成信息共享机制。教师网络任务特点是可利用的数据量大、规律性复杂,但学到信息的充分性和任务的聚焦性不高,学生模型的特点是利用标记信息直接作为监督信号,聚焦任务。在本文中,教师网络建模了无标记的时空动态特征以实现规律性建模,应对事件标记稀疏挑战、提升无标记数据利用率,而学生网络直接学习事件级联影响和特征-事件映射关系,聚焦任务本身,最终以师生配合的方式得到细粒度风险估计与离散的高风险区域提名。本文可视作半监督学习在时空离散数据预测领域的应用创新。

对于未来的研究方向,一方面,我们将事件预测任务与时空点过程结合,更精细化地建模事件影响的时空模式;另一方面,由于城市离散事件时空稀疏、不确定性大,我们将进一步探究离散事件的可预测性,并在事件预测方法中引入不确定性理论,为可靠城市计算提供关键算法支撑。

参考文献

- [1] ZHAO X Y, TANG J L. Exploring transfer learning for crime prediction[C]//2017 IEEE International Conference on Data Mining Workshops. Piscataway: IEEE, 2017: 1158-1159.
- [2] WANG B, LUO X Y, ZHANG F B, et al. Graph-based deep modeling and real time forecasting of sparse spatio-temporal data[EB/OL]. (2018-04-02)[2021-11-24]. <https://arxiv.org/abs/1804.00684>.
- [3] NIKOLOPOULOS K, PUNIA S, SCHÄFERS A, et al. Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions[J]. European Journal of Operational Research, 2021, 290(1): 99-115.
- [4] CHANG C Y, LU C W, WANG C J. A multi-step-ahead markov conditional forward model with cube perturbations for extreme weather forecasting[C]//Proc of the 35th AAAI Conference on Artificial Intelligence. California: AAAI, 2021: 6948-6955.
- [5] ZHOU Z Y, WANG Y, XIE X K, et al. Foresee urban sparse traffic accidents: A spatiotemporal multi-granularity perspective[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(8): 3786-3799.
- [6] ZHOU Z Y, WANG Y, XIE X K, et al. RiskOracle: A minute-level citywide traffic accident forecasting framework [C]//Proc of the 34th the Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence. California: AAAI, 2020: 1258-1265.
- [7] MOOSAVI S, SAMAVATIAN M H, PARTHASARTHY S, et al. Accident risk prediction based on heterogeneous sparse data: New dataset and insights[C]//Proc of the 27th ACM Sigspatial International Conference on Advances in Geographic Information Systems. New York: ACM, 2019: 33-42.
- [8] CHEN L R, GUAN Q F, XIONG Y H, et al. A spatially constrained multi-autoencoder approach for multivariate

- geochemical anomaly recognition[J]. *Computers & Geosciences*, 2019, 125: 43-54.
- [9] 黄志清, 曲志伟, 张吉, 等. 基于深度强化学习的端到端无人驾驶决策[J]. *电子学报*, 2020, 48(9): 1711-1719.
HUANG Z Q, QU Z W, ZHANG J, et al. End-to-end autonomous driving decision based on deep reinforcement learning[J]. *Acta Electronica Sinica*, 2020, 48(9): 1711-1719. (in Chinese)
- [10] 张啸剑, 孟小峰. 面向数据发布和分析的差分隐私保护[J]. *计算机学报*, 2014, 37(4): 927-949.
ZHANG X J, MENG X F. Differential privacy in data publication and analysis. *Chinese Journal of Computers*, 2014, 37(4): 927-949. (in Chinese)
- [11] YE Q Q, HU H B, AU M H, et al. LF-GDPR: A framework for estimating graph metrics with local differential privacy[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34(10): 4905-4920.
- [12] ZHAO J W, GAO Y J, MA C Y, et al. On efficiently diversified top-k geo-social keyword query processing in road networks[J]. *Information Sciences*, 2020, 512: 813-829.
- [13] GNEITING T, RAFTERY A E. Weather forecasting with ensemble methods[J]. *Science*, 2005, 310(5746): 248-249.
- [14] HOLMSTROM M, LIU D, VO C. Machine learning applied to weather forecasting[J]. *Meteorol*, 2016, 4: 1-5.
- [15] 冯宁, 郭晟楠, 宋超, 等. 面向交通流量预测的多组件时空图卷积网络[J]. *软件学报*, 2019, 30(3): 759-769.
FENG N, GUO S N, SONG C, et al. Multi-component spatial-temporal graph convolution networks for traffic flow forecasting[J]. *Journal of Software*, 2019, 30(3): 759-769. (in Chinese)
- [16] 杜圣东, 李天瑞, 杨燕, 等. 一种基于序列到序列时空注意力学习的交通流预测模型[J]. *计算机研究与发展*, 2020, 57(8): 1715-1728.
DU S D, LI T R, YANG Y, et al. A sequence-to-sequence spatial-temporal attention learning model for urban traffic flow prediction[J]. *Research and Development of Computers*, 2020, 57(8): 1715-1728. (in Chinese)
- [17] 丁治明, 韩京宇, 李曼, 等. 基于网络受限移动对象数据库的交通流统计分析模型[J]. *计算机研究与发展*, 2008, 45(4): 646-655.
DING Z M, HAN J Y, LI M, et al. Network-constrained moving objects database based traffic flow statistical analysis model[J]. *Research and Development of Computers*, 2008, 45(4): 646-655. (in Chinese)
- [18] ZHANG J B, ZHENG Y, QI D K. Deep spatio-temporal residual networks for citywide crowd flows prediction [C]//Proc of the 31st AAAI Conference on Artificial Intelligence. California: AAAI, 2017: 1655-1661.
- [19] HAO Z K, LU C Q, HUANG Z Y, et al. ASGN: An active semi-supervised graph neural network for molecular property prediction[C]//Proc of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 731-752.
- [20] YU B, YIN H T, ZHU Z X. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting[C]//Proc of the 27th International Joint Conference on Artificial Intelligence. Melbourne: IJCAI, 2018: 3634-3640.
- [21] BAI L, YAO L N, KANHERE S S, et al. STG2seq: Spatial-temporal graph to sequence model for multi-step passenger demand forecasting[C]//Proc of the 28th International Joint Conference on Artificial Intelligence. Melbourne: IJCAI, 2019: 1981-1987.
- [22] SONG C, LIN Y F, GUO S N, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting[C]//Proc of the 34th AAAI Conference on Artificial Intelligence. California: AAAI, 2020: 914-921.
- [23] 徐薇, 黄厚宽, 王英杰. 一种基于数据融合和方法融合的时空综合预测算法[J]. *计算机研究与发展*, 2005(7): 1255-1260.
XU W, HUANG H K, WANG Y J. An integrated spatio-temporal forecasting approach based on data fusion and method fusion[J]. *Research and Development of Computers*, 2005(7): 1255-1260. (in Chinese)
- [24] MA J, LI Z, JACK C P C, et al. Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network[J]. *Science of The Total Environment*, 2020, 705: 135771.
- [25] YI X W, ZHANG J B, WANG Z Y, et al. Deep distributed fusion network for air quality prediction[C]//Proc of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 965-973.
- [26] YUAN Z N, ZHOU X, YANG T B. Hetero-convLSTM: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data[C]//Proc of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 984-992.
- [27] FENG J, LIN Z Q, XIA T, et al. A sequential convolution

network for population flow prediction with explicitly correlation modelling[C]//Proc of the 29th International Joint Conference on Artificial Intelligence. Melbourne: IJCAI, 2020: 1331-1337.

- [28] SHI H Z, YAO Q M, GUO Q, et al. Predicting origin-destination flow via multi-perspective graph convolutional network[C]//Proc of the 36th IEEE International Conference on Data Engineering. Piscataway: IEEE, 2020: 1818-1821.
- [29] XU F L, LI Y, XU S S. Attentional multi-graph convolutional network for regional economy prediction with open migration data[C]//Proc of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 2225-2233.
- [30] ZHANG M Y, LI T, LI Y, et al. Multi-view joint graph representation learning for urban region embedding[C]//Proc of the 29th International Joint Conference on Artificial Intelligence. Melbourne: IJCAI, 2020: 4431-4437.
- [31] FENG J, LI Y, ZHANG C, et al. Deepmove: Predicting human mobility with attentional recurrent networks[C]//Proc of the 27th the Web Conference. New York: ACM, 2018: 1459-1468.
- [32] TENNESSEE M. Tennessee highway patrol provides predictive analytics software to all tennessee sheriffs' offices [EB/OL]. (2017-10-10)[2021-08-01]. <https://www.tn.gov/safety/news/2017/10/10/>.
- [33] CHANG L Y, CHEN W C. Data mining of tree-based models to analyze freeway accident frequency[J]. Journal of Safety Research, 2005, 36(4): 365-375.
- [34] HUANG C, ZHANG C X, DAI P, et al. Deep dynamic fusion network for traffic accident forecasting[C]//Proc of the 28th ACM International Conference on Information and Knowledge Management. New York: ACM, 2019: 2673-2681.
- [35] REN H L, SONG Y, WANG J W, et al. A deep learning approach to the citywide traffic accident risk prediction[C]//Proc of the 21st International Conference on Intelligent Transportation Systems. Piscataway: IEEE, 2018: 3346-3351.
- [36] CHEN C, FAN X L, ZHENG C P, et al. SDCAE: Stack denoising convolutional autoencoder model for accident risk prediction via traffic big data[C]//Proc of the 6th International Conference on Advanced Cloud and Big Data (CBD). Piscataway: IEEE, 2018: 328-333.
- [37] ZHOU Z Y. Attention based stack resnet for citywide traffic accident prediction[C]//Proc of the 20th IEEE International Conference on Mobile Data Management. Piscataway: IEEE, 2019: 369-370.
- [38] BAO J, LIU P, UKKUSURI S V. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data[J]. Accident Analysis & Prevention, 2019, 122: 239-254.
- [39] HONG L, HONG L, ZHENG Y, et al. Detecting urban black holes based on human mobility data[C]//Proc of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM, 2015: 1-10.
- [40] YAO H X, TANG X F, WEI H, et al. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction[C]//Proc of the 33rd AAAI Conference on Artificial Intelligence. California: AAAI, 2019: 5668-5675.
- [41] 徐兴荣, 刘聪, 李婷, 等. 基于双向准循环神经网络和注意力机制的业务流程剩余时间预测方法[J]. 电子学报, 2022, 50(8): 1975-1984.
- XU X R, LIU C, LI T, et al. Business process remaining time prediction: An approach based on bidirectional quasi-recurrent neural network with attention[J]. Acta Electronica Sinica, 2022, 50(8): 1975-1984. (in Chinese)
- [42] 许森, 刘宏飞, 苏岳龙. 考虑交通事件影响的城市道路行程时间预测[J]. 中国公路学报, 2021, 34(12): 229-238.
- XU M, LIU H F, SU Y L. Urban road travel time prediction considering impact of traffic even[J]. China Journal of Highway and Transport, 2021, 34(12): 229-238. (in Chinese)
- [43] 刘浩阳, 林耀进, 刘景华, 等. 由粗到细的分层特征选择[J]. 电子学报, 2022, 50(11): 2778-2789.
- LIU H Y, LIN Y J, LIU J H, et al. Hierarchical feature selection from coarse to fine[J]. Acta Electronica Sinica, 2022, 50(11): 2778-2789. (in Chinese)

作者简介



周正阳 男, 1995年9月出生, 江苏泰州人. 2018年在江南大学获得学士学位, 2023年在中国科学技术大学计算机科学与技术学院获得博士学位. 现为中国科大软件学院副研究员. 主要研究方向为时空数据挖掘、交通预测、城市事件预测等.

E-mail: zzy0929@mail.ustc.edu.cn



汪 炆(通讯作者) 男, 1980年4月出生, 安徽歙县人. 现任中国科学技术大学软件学院副院长、副教授、博士生导师. 主要研究方向为时空数据挖掘、智能交通系统及数据驱动的材料科学研究.

E-mail: angyan@ustc.edu.cn