



面向稀疏感知的城市 时空数据挖掘



中国科学技术大学 计算机科学与技术学院

中国科大-数据智能实验室 周正阳

导师 汪 炆



目 录

- 1 报告背景与介绍
- 2 时空数据挖掘与稀疏感知概述
- 3 主要技术路线
- 4 研究与应用（三个典型工作）
- 5 前沿研究

报告背景与报告人介绍

智慧城市与时空数据挖掘：随着新兴物联网技术的快速发展，可获得的城市数据呈指数级增长，大数据时代已经来临。爆发式增长的数据和人工智能技术使得数据赋能城市发展。



报告背景与报告人介绍

《交通强国建设纲要》（2019.09）

总体要求：

建设交通强国是以习近平同志为核心的党中央立足国情、着眼全局、面向未来作出的重大战略决策，是建设现代化经济体系的先行领域，是全面建成社会主义现代化强国的重要支撑，是新时代做好交通工作的总抓手。为统筹推进交通强国建设，制定本纲要。

五、科技创新富有活力、智慧引领

（二）大力发展智慧交通。推动大数据、互联网、人工智能、区块链、超级计算等新技术与交通行业深度融合。推进数据资源赋能交通发展，加速交通基础设施网、运输服务网、能源网与信息网络融合发展，构建泛在先进的交通信息基础设施。构建综合交通大数据中心体系，深化交通公共服务和电子政务发展。推进北斗卫星导航系统应用。

十九届五中全会精神

- 国际环境的要求：新一轮科技革命和产业变革深入发展，以自主创新为驱动，发展科技核心竞争力；
- 传统产业深度变革：推动传统产业与现代产业和信息技术深度融合；
- 城市数字化治理：科技强国提升城市治理效率和能力。

报告背景与报告人介绍

《交通强国建设纲要》（2019.09）

总体要求：

建设交通强国是以习近平同志为核心的党中央立足国情、着眼全局、面向未来作出的重大战略决策，是建设现代化经济体系的先行领域，是全面建成社会主义现代化强国的重要支撑，是新时代做好交通工作的总抓手。为统筹推进交通强国建设，制定本纲要。

五、科技创新富有活力、智慧引领

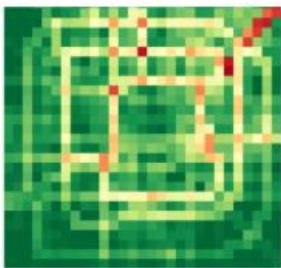
（二）大力发展智慧交通。推动大数据、互联网、人工智能、区块链、超级计算等新技术与交通行业深度融合。推进数据资源赋能交通发展，加速交通基础设施网、运输服务网、能源网与信息网络融合发展，构建泛在先进的交通信息基础设施。构建综合交通大数据中心体系，深化交通公共服务和电子政务发展。推进北斗卫星导航系统应用。

数据赋能城市发展是

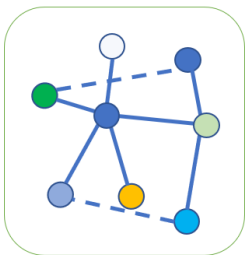
新时代的要求与使命

时空数据概述

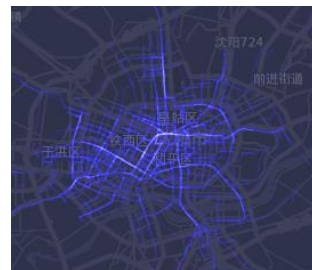
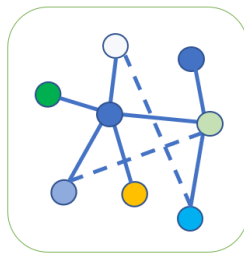
- **时空数据**是指具有空间分布且随着时间变化而变化的数据的集合。e.g. 路段（路口）交通流、人口密度、大气污染、网约车订单、国民经济发展、兴趣点（POI）签到。
- **来源**：遥感（RS）、城市监控（Monitoring）、移动设备（Mobile device）
- **数据形态**



网格型 (Grid)



图型 (Graph)



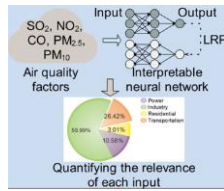
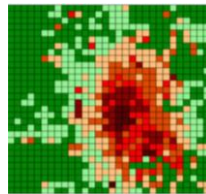
轨迹行为序列 (Trajectory)

时空数据概述

时空数据挖掘与城市计算

随着城市化发展与城市人口增长，人们的生产、出行行为不断影响着周围的社会自然环境，**交通拥堵、房价暴涨、大气污染**愈加困扰着人们的生活。大数据与人工智能时代的到来，似乎又为这些恼人的问题带来了一线希望。基于**多源异构的时空数据挖掘与稀疏感知**技术，可深入发掘潜在关联，揭示人类活动与自然、社会的关系，从而制定政策、指导性意见更好的规范人的行为，促使城市可持续发展。

当城市计算遇上人工智能 Urban Computing meets AI



时空数据概述

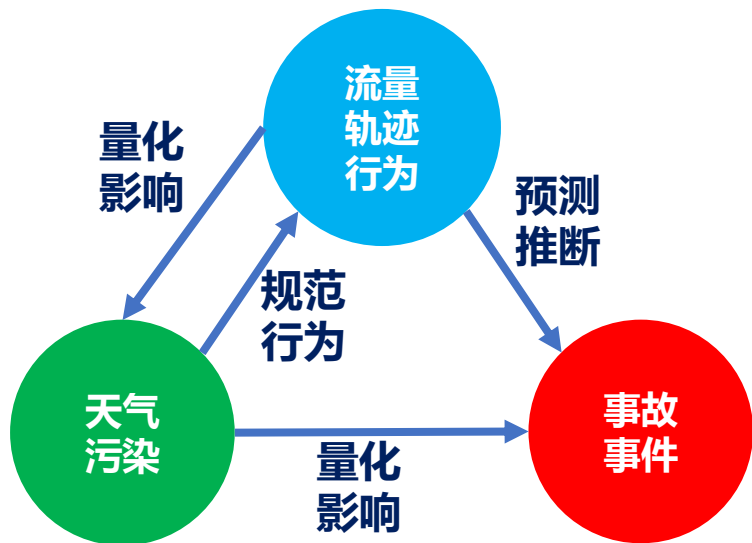
时空数据挖掘的价值和意义

时空数据主要包含了人类活动与气象天气等可量化数值的时空分布，恰好映射了社会经济环境与自然环境两个方面。



时空数据概述

时空数据挖掘的价值和意义



- ★ 理解人类活动与时空语义信息的潜在关联
- ★ 以数据视角服务城市可持续发展

时空数据概述

时空数据挖掘的科学问题与主要任务

• 科学问题

多源异质信息协同建模

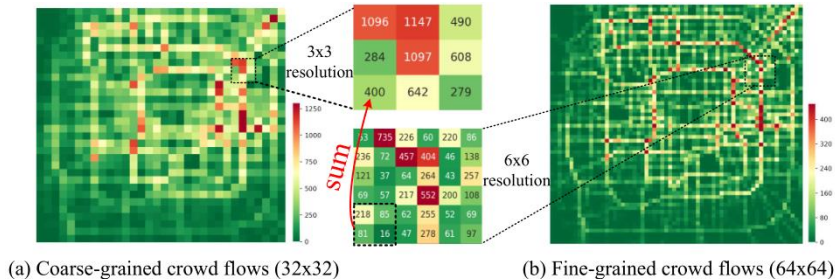
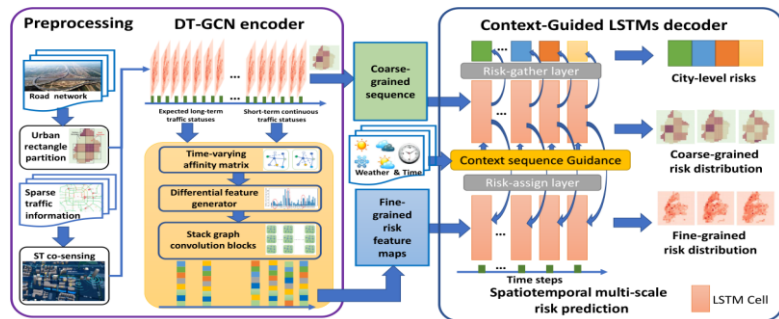
多源时空数据关联性解耦

• 主要任务

时空预测问题：基于历史数据预测未来

时空推断问题：基于历史数据及多源实时数据进行全局推断

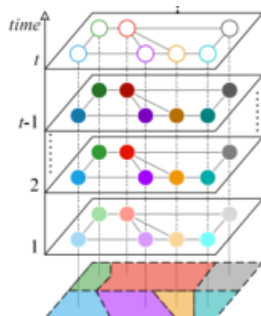
时空关联性解析：对两者的潜在关联性进行研究



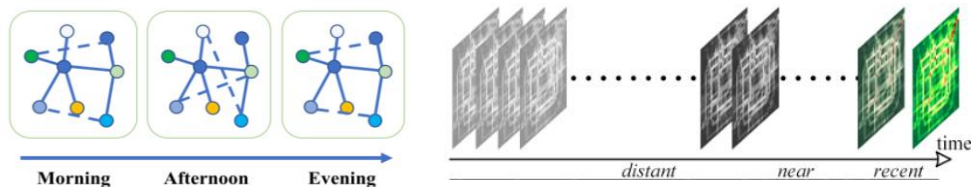
时空数据概述

时空问题特点 (交通为例)

1. 路网传播特性



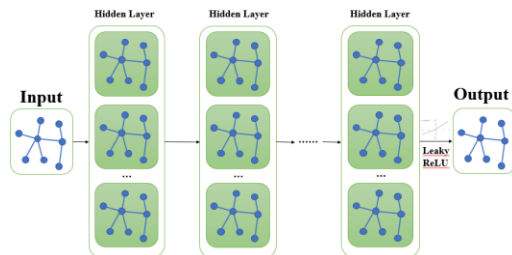
2. 区域之间的动态空间关联



3. 时间序列的趋势性、连续型、周期性

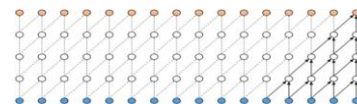
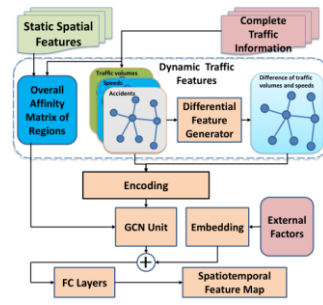
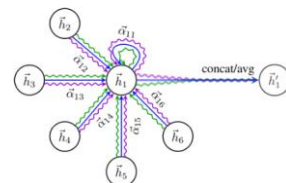
针对性技术

网格数据CNN、图网络建模GCN



注意力机制、GAT

抽取时变关联



RNN变体时间序列建模、TCN

技术路线

时空数据挖掘的主要研究方法与技术

- 将空间建模为网格或图

引入计算机视觉相关建模技术 (ResNet, DenseNet, ST-GCN) 。

- 主数据集+辅助数据集

基于辅助数据集的协同推断，挖掘不同的映射函数以建模时空关联性。

- 基于半监督的推断学习

通过学习节点的特征表示映射，从而对未在训练集中的节点直接生成表示从而进行推断（分类）等。

技术路线

城市时空数据挖掘问题分类

- 基本元素的推断与预测（房价、速度、流量、污染量等）
- 离散事件的推断与预测（打车需求、路网事故、拥堵事件）
 - 挖掘基本元素与人类活动之间的潜在关联；
 - 离散事件本身的时间序列建模；

挑战： 离散/零星/稀疏的历史数据；
人类行为随机性影响。

技术路线

时空数据挖掘一般技术路线

传统机器学习

事件预测 (Classification)

Xgboost, Random Forest

流量速度预测 (Regression)

非负矩阵分解NMF

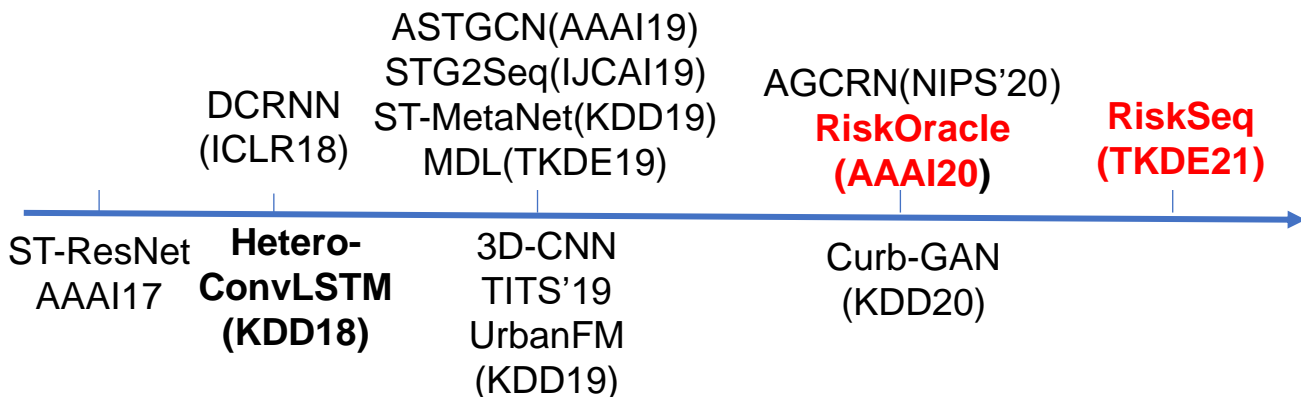
核密度估计 KDE

滑动平均自回归ARIMA

Graph-based

深度学习

Grid-based



问题与挑战

时空数据中的两类稀疏挑战

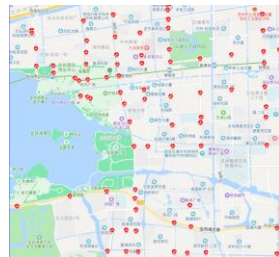
本质稀疏： 本身具有少量标签且难以生成新的有效标签的问题叫做本质稀疏，如犯罪事件、交通拥堵（事故）事件、偶发疫情。**无法补全。**

伪稀疏： 数据本身存在然而却因为外界原因未被感知或探测（未完整感知）如城市的速度 loop detector、气象/大气环境数据等。**可适当填充。**

- 训练过程的零膨胀问题（zero-inflated issue）
- 有效数据覆盖面小，难以支撑训练



(a) 本质稀疏



(b) 伪稀疏



面向稀疏感知的城市时空数据挖掘

- (1) 全城多粒度稀疏交通事故预测
- (2) 未来感知与上下文感知的城市出租车调度
- (3) 基于大规模稀疏轨迹的细粒度轨迹预测

全城多粒度稀疏交通事故预测

◆ 研究背景与价值意义

- 提升公共资源分配的精准性与公平性
- 维护城市公共安全，提升幸福感和安全感

◆ 相关工作

【长期预测】针对全州（全省）下一周中**每日事故总数**进行预测，形成事故风险图，但均为固定的空间尺度。

【短期预测】1h的时间粒度，且均为**单步预测**，不能感知到路网的短期变化，未解决**本质稀疏**带来的零膨胀问题。

◆ 空间多尺度与短期多步预测

- 不同级别交警部门的多样化需求
- 市民出行规划和城市管理预警



Captured information and traffic volume analysis of a camera



Citywide traffic surveillance and accident spots



TKDE'21: Foresee Urban Sparse Traffic Accidents:
A Spatiotemporal Multi-Granularity Perspective

全城多粒度稀疏交通事故预测

◆ 研究背景与价值意义

- 提升公共资源分配的精准性与公平性
- 维护城市公共安全，提升幸福感和安全感

◆ 相关工作

【长期预测】针对全州（全省）下一周中**每日事故总数**进行预测，形成事故风险图，但均为固定的空间尺度。

【短期预测】1h的时间粒度，且均为**单步预测**，不能感知到路网的短期变化，未解决**本质稀疏**带来的零膨胀问题。

问题：时空多尺度交通事故预测：给定静态路网结构特征 \mathcal{S} 和历史的动态交通信息 $\mathcal{F}(\Delta t)$ ($\Delta t = 1, 2, \dots, T$)，我们的任务是同时预测在未来 r 步空间上粗粒度和细粒度的交通事故风险，以及最可能发生事故风险的 M 个区域，即 $\mathcal{O}_C(\Delta t')$ ， $\mathcal{O}_F(\Delta t')$ ，和 $\mathcal{V}_M(\Delta t')$ ，其中 $\Delta t' = T + 1, T + 2, \dots, T + r$ 。



Captured information and traffic volume analysis of a camera



Citywide traffic surveillance and accident spots



全城多粒度稀疏交通事故预测

基于多源历史数据预测未来多粒度事故分布

数据预处理: PKDE & ST-DFM

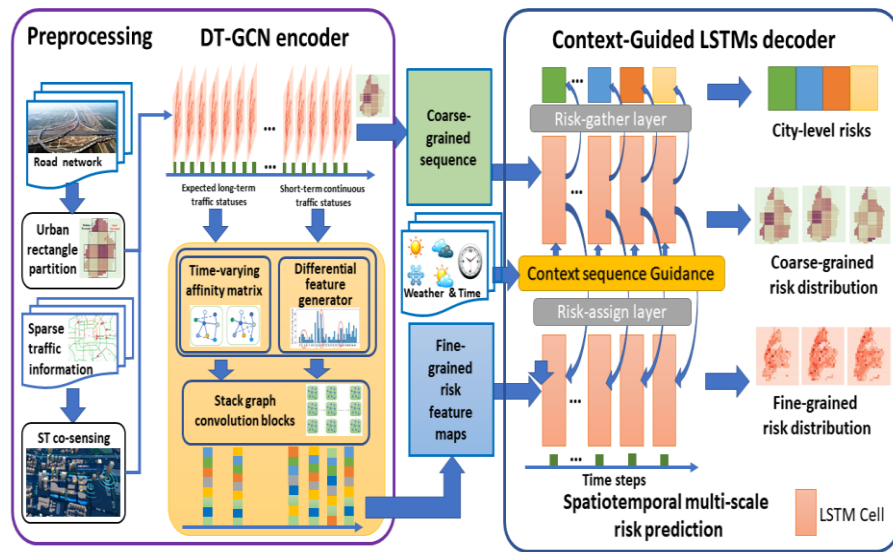
- 空间多粒度网格划分与稀疏动态数据实时推断
- 面向零膨胀问题的稀疏事故数据变换

时空建模: DTGCN & CG-LSTM

- 路网动态变化与短期状态异常变化捕获
- 上下文引导的多步预测与空间多尺度依赖建模

后处理-事故筛选

- 自适应事故区域Ranking机制



TKDE'21: Foresee Urban Sparse Traffic Accidents:
A Spatiotemporal Multi-Granularity Perspective



全城多粒度稀疏交通事故预测

数据预处理

(1) 层次性城市网格划分

(2) 基于数据集先验信息的本质稀疏缓解方法

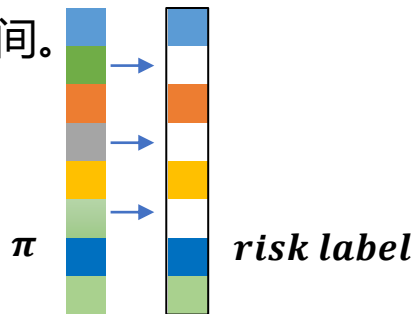
step1: 统计每个区域在这个数据集上的事故总数，并转换成一个0~1之间的概率值 ε_{v_i}

step2: 将 ε_{v_i} 利用对数log 转化成一个负数 π_{v_i} ，并且使用一些约束的参数b1、b2来使其和正的风险值相一致，如正的风险值在0-5之间，那么负数值也在-5~0之间。

Motivation

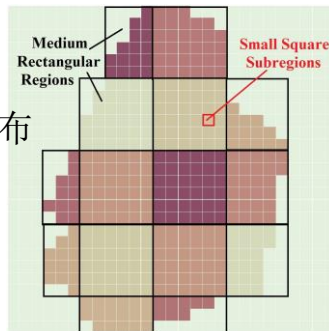
- ✓ 扩大正负样本距离；
- ✓ 显著区分潜在风险不同的区域；
- ✓ loss与label的一致性

$$\varepsilon_{v_i} = \frac{1}{N_{week}} \sum_{j=1}^{N_{week}} \frac{r_{v_i}(j)}{\sum_{k=1}^m r_{v_k}(j)} \quad \pi_{v_i} = b_1 \log_2 \varepsilon_{v_i} + b_2$$



Motivation

- ✓ 空间粒度降低，准确率提高
- ✓ 收集不同空间尺度的事故分布
- ✓ 便于设计事故区域提名策略





全城多粒度稀疏交通事故预测

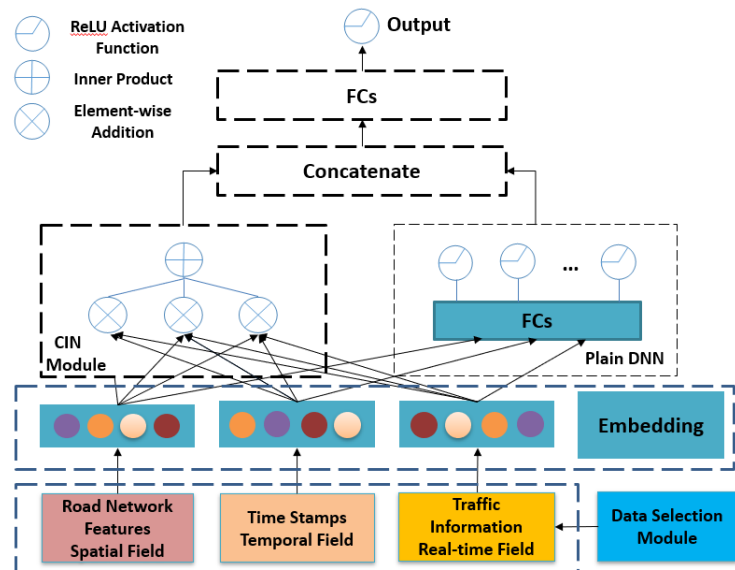
数据预处理

(3) 基于时空深度因子分解机的伪稀疏缓解策略

- 构造城市图 无向的城市图 $G(\mathcal{V}, \mathcal{E})$. 其中 $\mathcal{V} = \{v_1, v_2, v_i, \dots, v_m\}$

$$\alpha_s(i, j) = \begin{cases} 1 & \text{if subregion } v_i \text{ and } v_j \text{ are geographically adjacent} \\ e^{-JS(s_i \| s_j)} & \text{otherwise} \end{cases}$$

- 对每个区域而言，设计压缩交互网络CIN和DNN模块
- 筛选对应的静态路网特征，时间戳分别放入1,2 field
- 筛选最邻近的区域的动态信息放入Real-time field



Motivation

- ✓ 多源时空特征存在高阶交互影响，e.g. 路网结构与天气形成“共振”
- ✓ 车流与车速等动态特征存在非线性关联



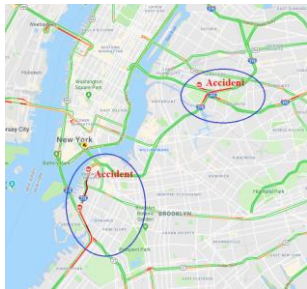
全城多粒度稀疏交通事故预测

时空建模：时变图卷积的神经网络 (Differential Time-varying Graph Convolution Network, DT-GCN)

GCN动机：

- (1) 由于交通事故和道路拥堵存在一定的交互影响和传播关系；
- (2) 相似的路网结构和相似的动态交通模式可能事故共现（非欧氏关联）；
- (3) GCN适合建模传播关系与非欧氏关联。

两个处于交叉路口的
地区发生拥堵和事故



相似的路网结构/
动态车流模式



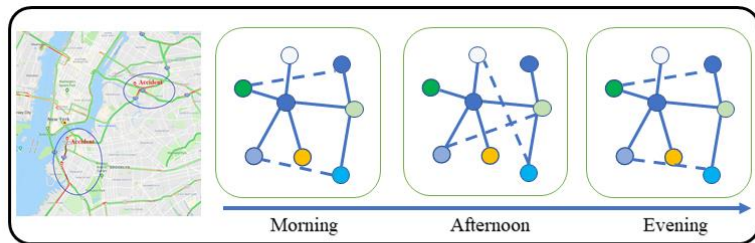
事故共现
Concurrence

全城多粒度稀疏交通事故预测

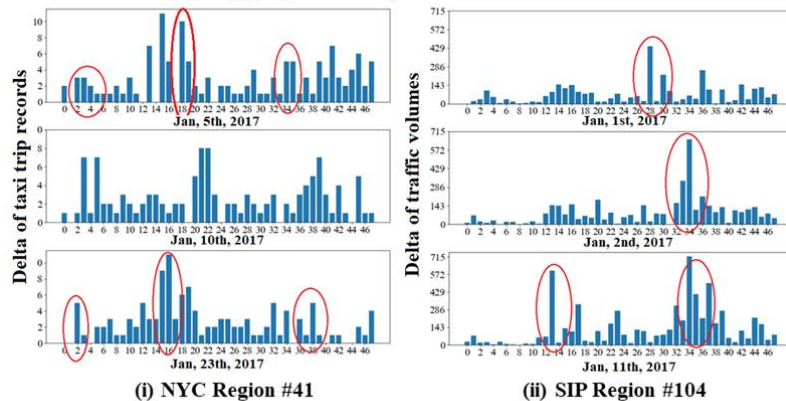
时空建模 (DT-GCN)

(1) 强烈的时变特性：区域之间交通模式存在一定的相似性和关联性，这种关联因潮汐车流等原因产生的会随时间变化的不同关联程度。

(2) 交通状态突变与事故关联：对于同一区域，相邻时间间隔内交通基础元素的数值变化对交通事故的影响（贡献）。



(a) Dynamic region-wise correlations



(b) Circled Accidents with regard to immediate changes of traffic volumes

全城多粒度稀疏交通事故预测

时空建模 (DT-GCN)

(1) 强烈的时变特性: 区域之间交通模式存在一定的相似性和关联性, 这种关联因潮汐车流等原因产生的会随时间变化的不同关联程度 -> 时变亲和度矩阵

Overall affinity matrix

$$\alpha_o^{\Delta t}(i, j) = e^{-JS(s_i^* \| s_j^*)} + \gamma * e^{-JS(C_i^{\Delta t} \| C_j^{\Delta t})} + \beta * tr_{ij}^{\Delta t}$$

静态特征相似性 动态特征相似性 区域i到区域j的
流量转移

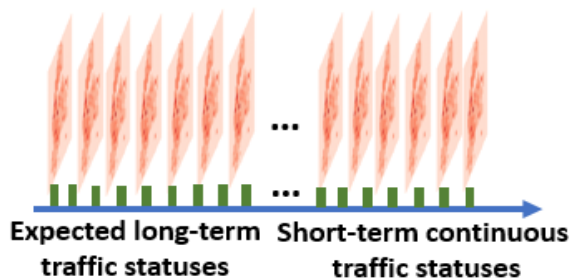
(2) 交通状态突变: 对于同一区域, 相邻时间间隔内交通基础元素的数值变化对交通事故的影响 (贡献) -> 差分特征

$$\vec{\Theta}^{\Delta t} = \mathcal{D}(\Delta t) - \mathcal{D}(\Delta t - 1)$$



全城多粒度稀疏交通事故预测

时空建模 (DT-GCN)



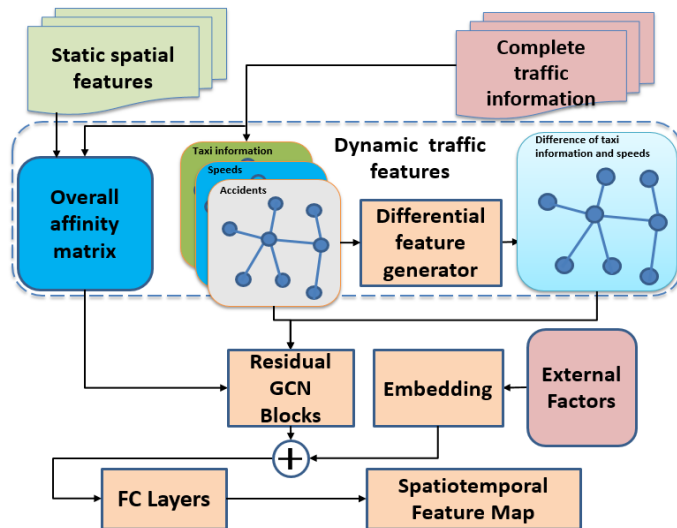
$$B^{\Delta t} = A_o^{\Delta t} + I_m$$

$$A_C^{\Delta t} = (\Phi^{\Delta t})^{-\frac{1}{2}} B^{\Delta t} (\Phi^{\Delta t})^{-\frac{1}{2}}$$

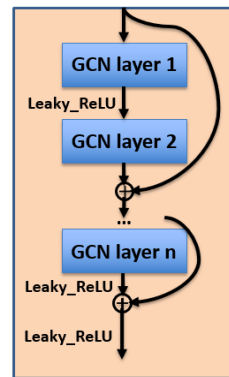
$$\mathcal{U}(\Delta t) = \left\{ \mathcal{F}(\Delta t), \vec{\Theta}^{\Delta t} \right\}$$

$$\mathcal{H}_{n+1} = \text{Leaky_ReLU}(A_C^{\Delta t} \mathcal{H}_n W_n)$$

where $\mathcal{H}_0 = \mathbb{U}_*^{\Delta t}$



(a) Architecture of DT-GCN



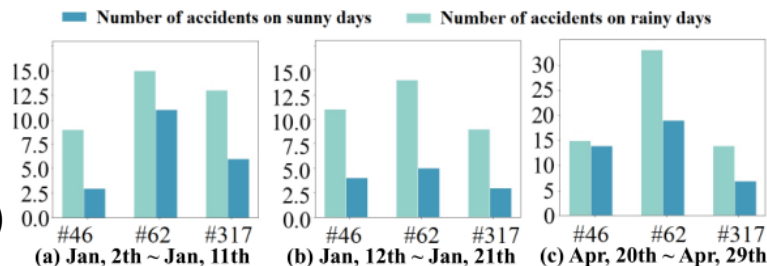
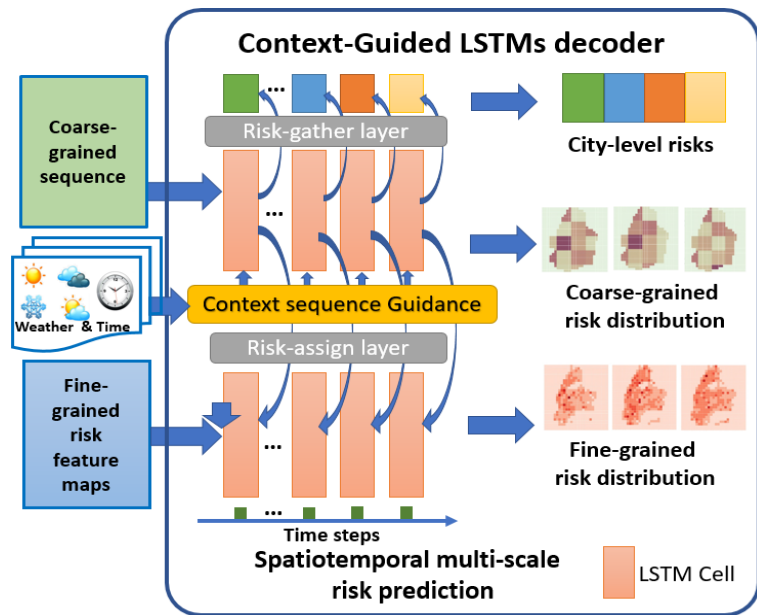
(b) Detailed Residual GCN block

$$\mathcal{M}_F = \{\mathcal{M}_F^0, \mathcal{M}_F^1, \dots, \mathcal{M}_F^{h+1}\}$$



全城多粒度稀疏交通事故预测

时空建模 (Context-Guided LSTM, CG-LSTM)



- ✓ 稀疏事件零膨胀问题：粗粒度风险序列易于预测；
- ✓ 空间异质：季节性影响，空间上对天气的敏感程度不同；
- ✓ 多步预测：不同时间步所产生事故风险的大小也不同。

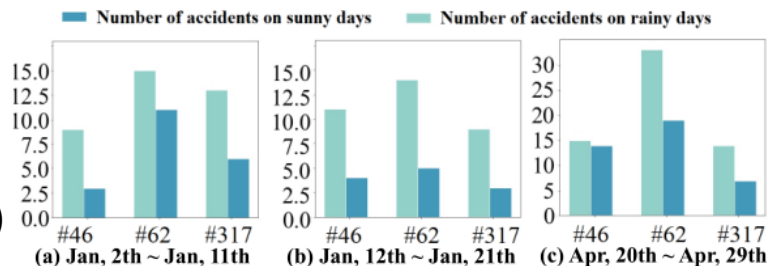
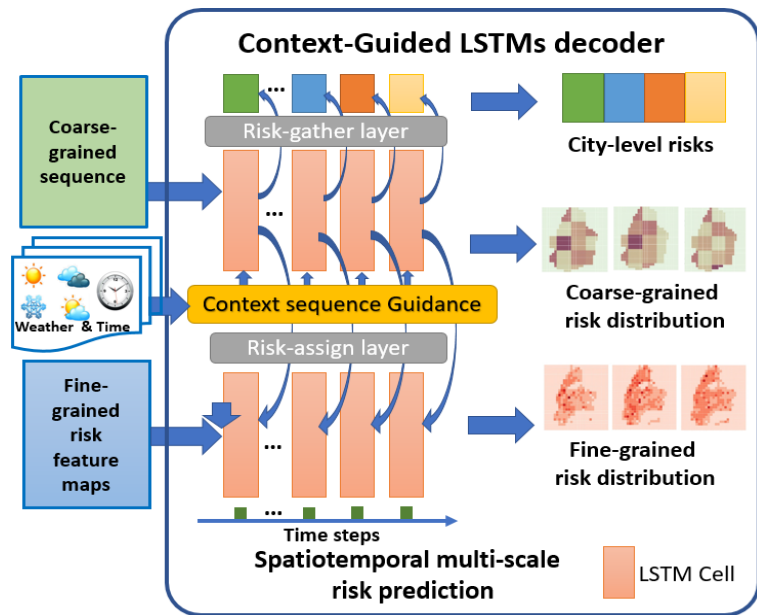


- ◆ 将粗粒度事故分布视为中间学习信息；
- ◆ 在解码器中引入上下文（天气、时间）；
- ◆ 风险空间多粒度依赖：Risk-gather/assign Layer 来聚合和分配风险。



全城多粒度稀疏交通事故预测

时空建模 (Context-Guided LSTM, CG-LSTM)



$$\mathcal{I}_C^{\Delta t+1} = \text{LSTM}_C(\mathcal{M}_C^{\Delta t+1}, [W_{\text{ext}} * E^{\Delta t+1} + \mathcal{I}_C^{\Delta t}])$$

$$\mathcal{I}_F^{\Delta t+1} = \text{LSTM}_F(\mathcal{M}_F^{\Delta t+1}, [W_{\text{asn}} * \mathcal{I}_C^{\Delta t} + \mathcal{I}_F^{\Delta t}])$$

City-level risk $\tilde{R}_S^{\Delta t} = W_{\text{gath}} * \mathcal{I}_C^{\Delta t}$

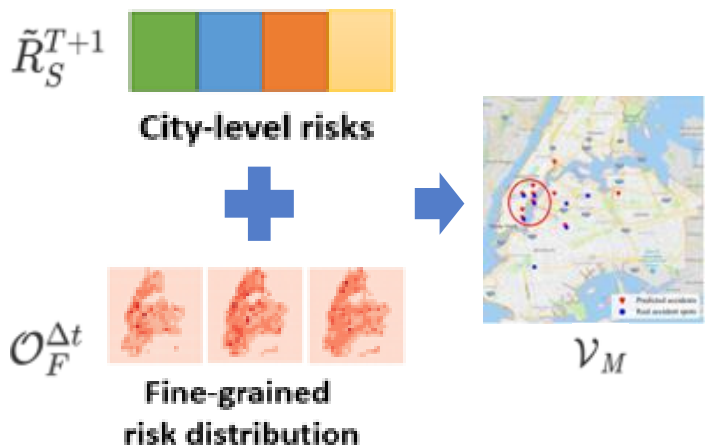
Coarse risk $\mathcal{O}_C^{\Delta t} = \text{ReLu}(W_{CF} * \mathcal{I}_C^{\Delta t} + b_{CF})$

Fine-grained risk $\mathcal{O}_F^{\Delta t} = \text{Leaky_ReLu}(W_{RF} * \mathcal{I}_F^{\Delta t} + b_{RF})$

$$\text{Loss}(\theta) = \text{MSE}_F + \lambda_1 * \text{MSE}_C + \lambda_2 * \text{MSE}_R + \lambda_3 * \text{L2}$$

全城多粒度稀疏交通事故预测

后处理阶段：自适应事故区域筛选



$$\{ \langle \mathcal{O}_F^{T+1}, \mathcal{O}_C^{T+1}, \tilde{R}_S^{T+1} \rangle, \dots, \langle \mathcal{O}_F^{T+r}, \mathcal{O}_C^{T+r}, \tilde{R}_S^{T+r} \rangle \}$$

每一时间步，将全城事故风险视为事故总数，自主地选择 $K(\Delta t) = \text{int}(\tilde{R}_S^{\Delta t})$ 选取个数阈值结合细粒度风险值，选取得到Top-K区域



全城多粒度稀疏交通事故预测

实验

数据集：NYC & SIP

City	Dataset ⁵	Time Span	# of Regions	# of Records
NYC	Accidents	01/01/2017- 05/31/2017	354	254k
	Taxi Trips			48,496k
	Speed Values			125k
	Weathers			604
	Demographics	Investigated in 2016		195
	Road Network			102k
SIP	Accidents	01/01/2017- 03/31/2017	108	183
	Traffic Flows			1,399k
	Speed Values			311k
	Weathers			180

评估指标

回归视角：MSE

分类视角：Acc@M 关注Hit到的准确率

选取风险最高的M个区域和真实发生事故区域进行比较

$$Acc@M = \frac{|V_{GT} \cap V_M|}{M}$$



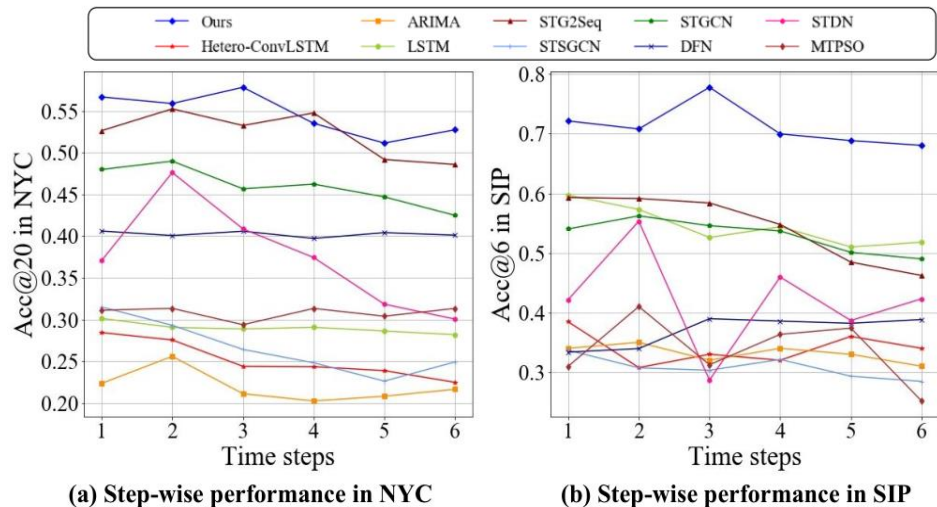
全城多粒度稀疏交通事故预测

实验

模型横向比较-平均结果

Models	NYC/SIP		
	Acc@20/ Acc@6	MSE-F	MSE-C
ARIMA	20.72/30.63	0.0192/ 0.0162	0.0492/0.2215
LSTM	28.98/35.70	0.0179/0.0255	0.0477/0.2694
Hetero-ConvLSTM	28.03/34.84	0.0161/0.0487	0.1015/0.4039
STGCN	50.42/51.27	0.0188 /0.0452	0.0492/0.2885
STG2Seq	52.08/54.30	0.0138/0.0364	0.0693/0.1667
STSGCN	26.46/33.59	0.0183/0.0236	0.1285/0.3473
STDN	37.48/42.18	0.0203 /0.0354	0.0853/0.2142
DFN	40.26/36.98	0.0194 /0.0376	0.0548/0.2278
MTPSO	30.81/33.69	0.0218 /0.0420	0.0393/0.2065
RiskSeq	56.42/71.27	0.0158/0.0401	0.0443/0.2702

模型横向比较-多步预测质量





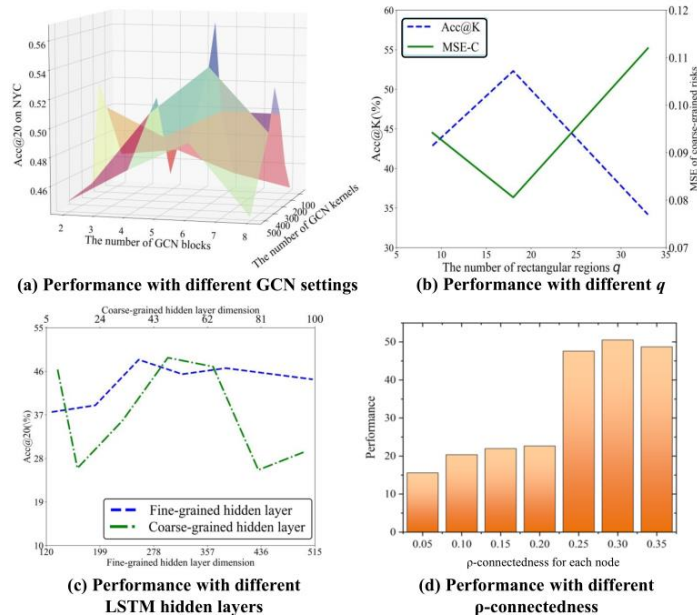
全城多粒度稀疏交通事故预测

实验

模型纵向比较-消融实验

Variant	MSE	NYC/SIP	
		Acc@20(Acc@6)	Acc@K
RS-PKDE	0.0053/0.0512	18.56/35.48	16.28/29.45
RS-DFM	0.1260/0.0216	43.05/58.94	38.29/46.28
RS-OA	0.0116/0.0127	37.57/67.16	32.47/61.15
RS-DG	0.0118/0.0136	46.45/68.52	39.19/55.27
RS-RC	0.0208/0.0082	41.79/69.45	38.19/56.33
RS-CF	0.0123/0.0355	43.04/67.83	33.21/50.18
RS-CGLSTM	0.0128/0.0060	48.45/67.19	-
Integrated RS	0.0158/0.0040	56.42/71.27	47.18/65.26

模型纵向比较-超参数网格搜索

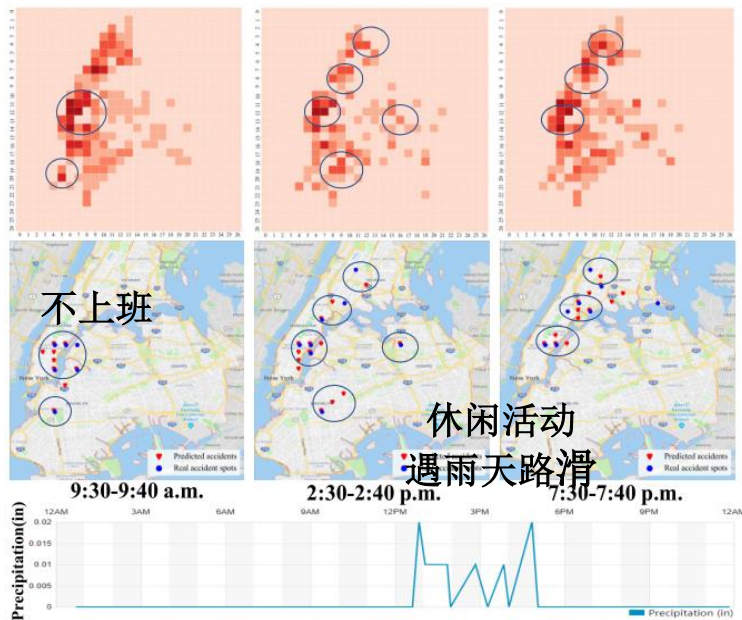


全城多粒度稀疏交通事故预测

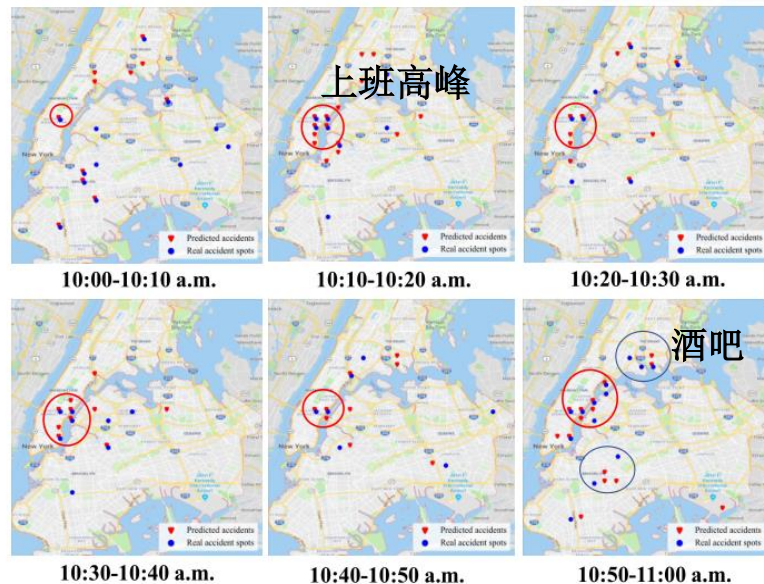
实验

案例分析:

- √ 预测事故风险分布大致吻合, 筛选结果较好
- √ 高风险区域呈现明显的时变特性
- √ 预测结果能够跟随上下文变化



(a) Three selected intervals on 22th, April, Sat, Cloudy to rainy



(b) Sequential results on 3rd, April, Mon, Rainy



全城多粒度稀疏交通事故预测

总结与讨论

RiskSeq通过**动态聚合邻域内的图信号**以获得更好的风险表示，并采用**逐步的上下文注入和多尺度的时间序列**学习来提升多步事故预测能力。

- **模型可扩展性**：犯罪和流行病预测、推荐系统物品推荐，**偶尔发生并表现出时变的空间依赖和人员流动（人类行为）模式**。进行数据处理后输入DT-GCN，捕获时变和异常情况，并使用CG-LSTM增强多步预测，可实现多粒度的城市事件预测。
- **稀疏时空数据挖掘**：**从稀疏性起源的角度**，缓解本质稀疏和伪稀疏问题，将稀疏事件预测转换为可通过DNN解决的可学习的回归和排序任务。

全城多粒度稀疏交通事故预测

总结与讨论

➤ 深度学习时空建模

- 转化思想：将稀疏事件转换为DNN可学习的回归与排序问题
- 网络结构：图中Edge时变设计结构、图信号signal-wise操作
- 稀疏感知：对问题进行分类解决，以分类的视角设计解决方案
- 困难学习问题：引导机制、多源信息的迁移与半监督学习
- 全局信息与局部信息：车 (local) 路(global) 协同

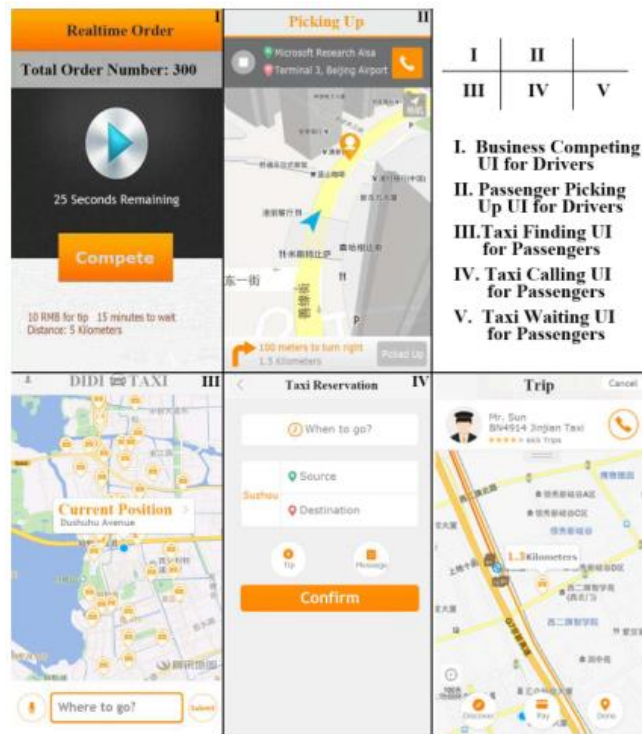
未来感知与上下文感知的城市出租车调度

大规模城市出租车调度：订单匹配+路径规划

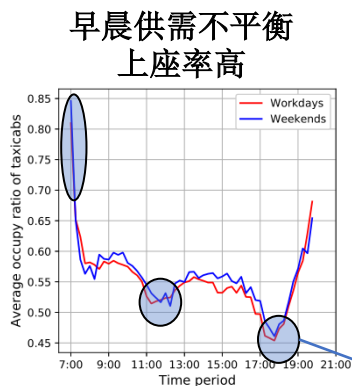
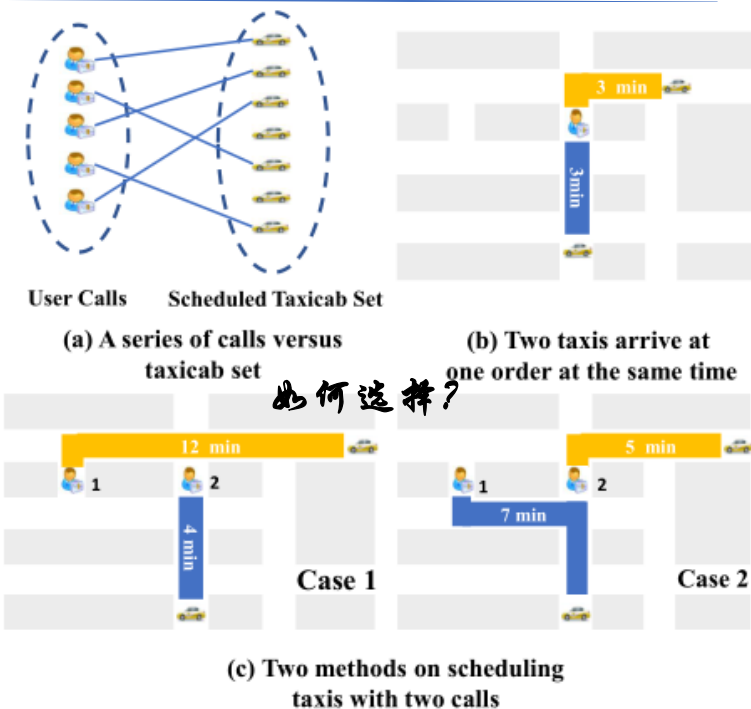
给定历史出租车轨迹 \mathcal{TR} 和对应车辆及其订单记录 \mathcal{A} ，我们需要学习订单分配与调度函数 $F(\text{order}, \text{driver})$ ，同时优化订单成交率 $R(d)$ 和总利润 $f(d)$ ：

$$\{\mathcal{TR}, \mathcal{A}\} \xrightarrow{F(\text{order}, \text{driver})} \text{Optm}\{R(d), f_p(d)\}$$

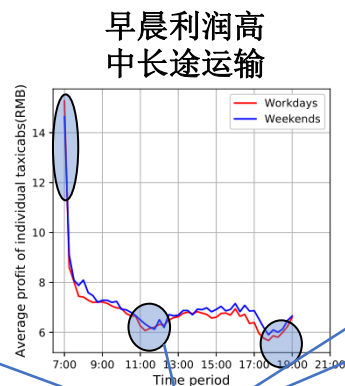
Large-Scale Intelligent Taxicab Scheduling: A Distributed and Future-Aware Approach (IEEE TVT 2020).



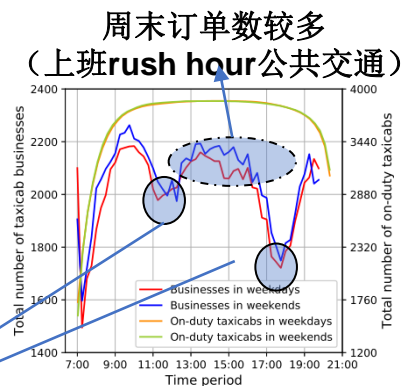
未来感知与上下文感知的城市出租车调度



Average occupied ratio



Average profit



Taxicab business

共同特点:

(1) 交通的谷期在饭点

(2) 周末倾向于下午出行, 下午订单和上座率均较高



未来感知与上下文感知的城市出租车调度

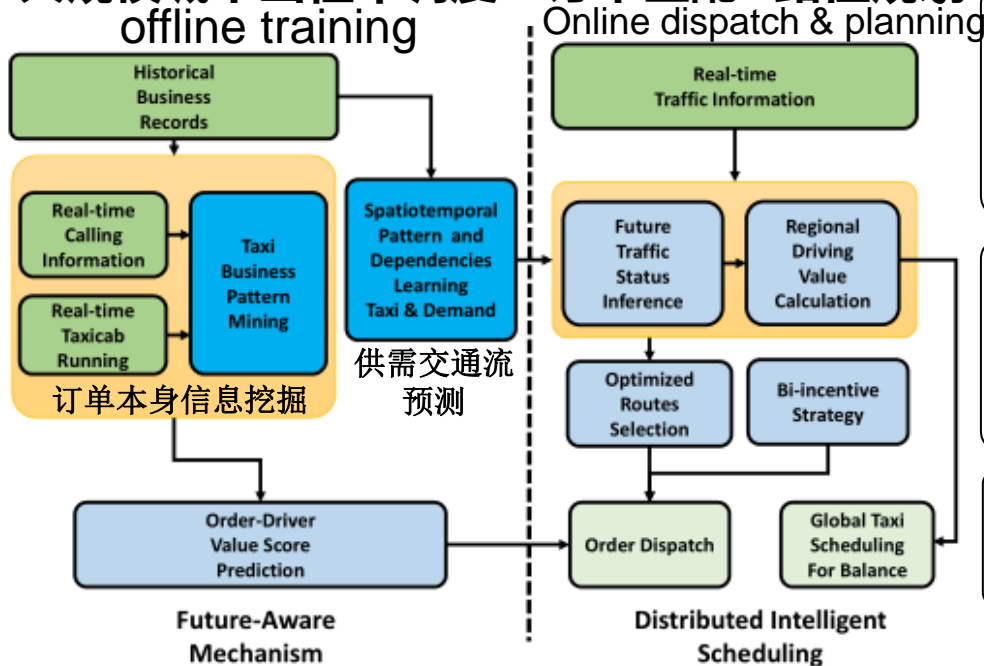
Motivation	Observation	Method
多数模型缺乏对 未来感知 预测	未来可预测，空车与叫车订单存在时空关联	(1) 未来感知的出租车需求、空车预测
多为贪心算法， 目标短视	价值分布由供需决定	(2) 区域驾驶价值评估RDV
当前订单 接受率与 全局 成交率难以平衡		(3) 订单-司机 交互价值预测
订单匹配与供需不平衡仍具挑战	订单接受概率受用户-司机历史偏好影响并决定于实时状态	供需不平衡下： 双边激励策略 (Bi-incentive)

Large-Scale Intelligent Taxicab Scheduling: A Distributed and Future-Aware Approach (IEEE TVT 2020).



未来感知与上下文感知的城市出租车调度

大规模城市出租车调度：订单匹配+路径规划



Future-aware Prediction

$$Ex[Tx(TP)] = \sum_{i \in RT(TP)} TM * Tx(i) \odot W_i^T$$
$$Ex[Tx(TP)] = \sum_{i \in RT(TP)} TM * Tx(i) \odot W_i^T$$

Context-Aware 区域价值

$$\mathcal{V}_{RDV}(r_i, TP) = 1 + \frac{1 + D_{r_i}(TP)}{1 + Tx_{r_i}(TP)}$$
$$\mathcal{V}_{RDV}(S, TP) = \prod_{r_i \in S} \mathcal{V}_{RDV}(r_i, TP)$$

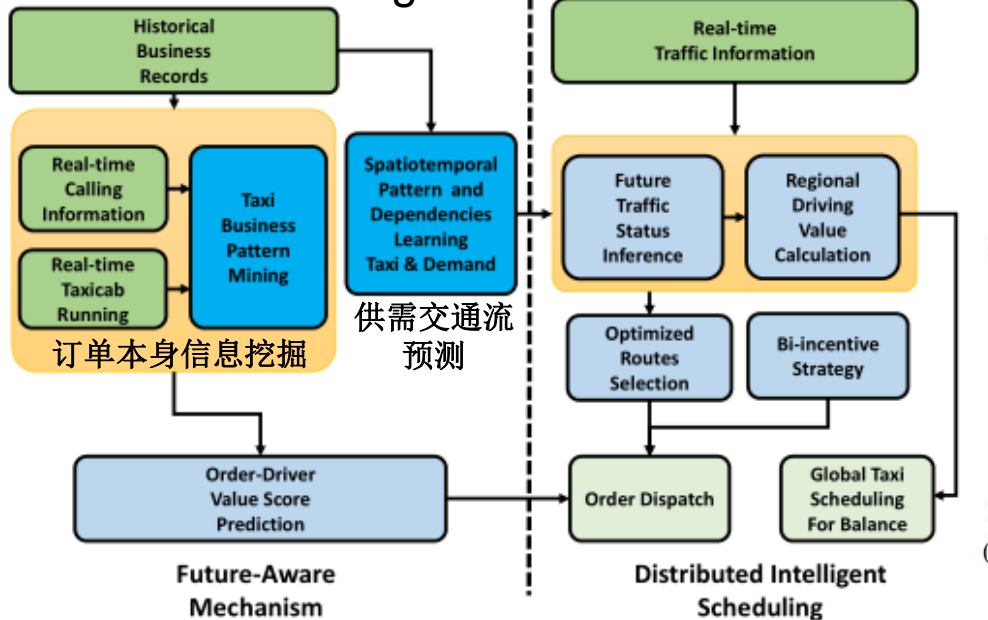
司机-订单匹配度预测

$$\mathcal{V}_{Ood} = p(y = 1 | r_k, r_o, r_d) = \frac{1}{\exp(-\mathbf{w}^T \mathbf{x}_{k,o,d})}$$

选取Top-K后再根据历史模式计算潜在收益

未来感知与上下文感知的城市出租车调度

大规模城市出租车调度：订单匹配+路径规划

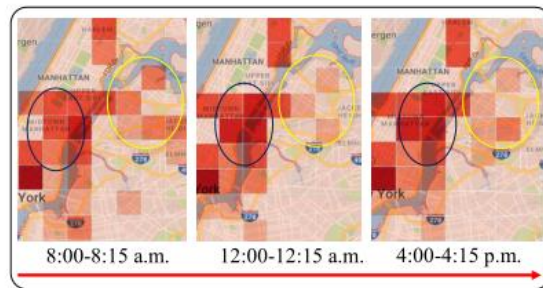


利润计算

订单匹配

分段
路径规划根据RDV和
供需预测结果

全局调度



(a) Time-varying RDV maps of a selected day in NYC



(b) Example of order dispatch



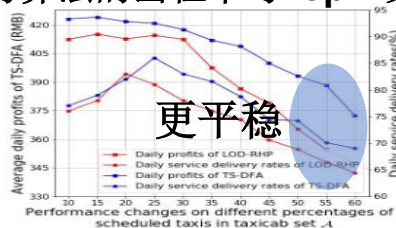
未来感知与上下文感知的城市出租车调度

大规模城市出租车调度：订单匹配+路径规划

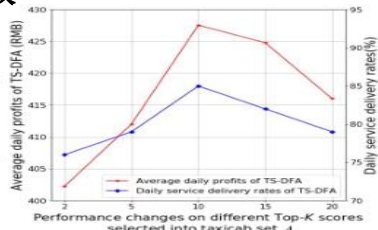
实验：NYC & SIP

City	Methods	Average Profits per day (RMB/USD)	Passenger Waiting Time (min)	Service Delivery Rates (%)
Suzhou	CERS	406.0	6.29	74.82
	TODCO	395.6	6.42	82.45
	LOD-RHP	418.5	5.69	81.03
	TS-DFA (Ours)	428.7	5.76	84.96
NYC	CERS	223.5	4.87	78.94
	TODCO	218.0	4.66	82.42
	LOD-RHP	238.3	4.32	80.36
	TS-DFA (Ours)	254.8	4.48	83.17

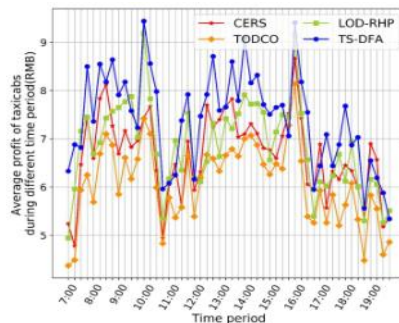
参与算法的出租车与Top-K数



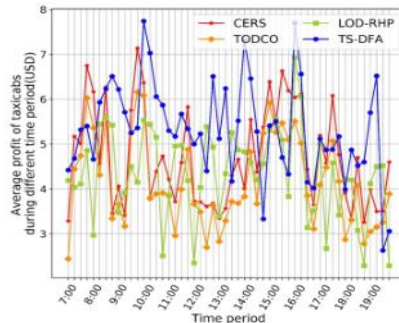
(a) Performance on different scheduled taxis



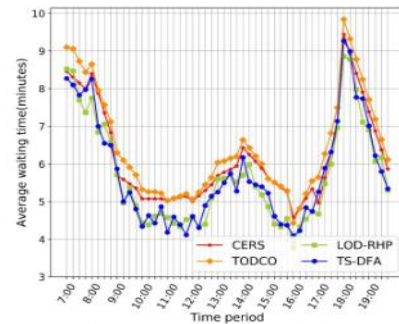
(b) Performance on different Top-K scores selected



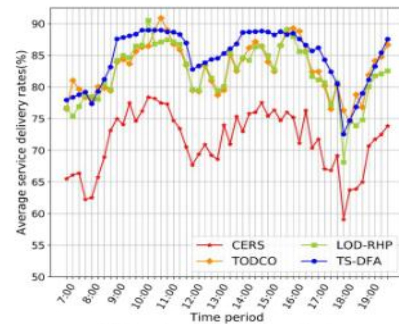
(a) Average profits during different time periods on Suzhou



(b) Average profits during different time periods on NYC



(c) Average passenger waiting time during different time periods on Suzhou



(d) Service delivery rates during different time periods on Suzhou

未来感知与上下文感知的城市出租车调度

大规模城市出租车调度：订单匹配+路径规划

讨论与思考：可学习的参数+传统的模式挖掘

(1) 当参与数量减少或历史可获得的匹配订单量远小于实际订单量时，如何利用少量信息来获得最大的算法效益？

(2) learnable K (司机-乘客匹配)，自适应地选择适合的candidate taxicabs.

(3) 深度强化学习的应用，action: route; reward: values & delivery rate (一致优化) ; state: current profits

大气科学与深度学习

气象环境特点

1. 空间传播特性:

乌云向邻近区域飘散 (降水)

大气扩散 (污染)

2. 时间邻近与周期特性

3. 人类活动相关

类比时空交通数据



时空数据挖掘的前沿研究

时空数据的安全性

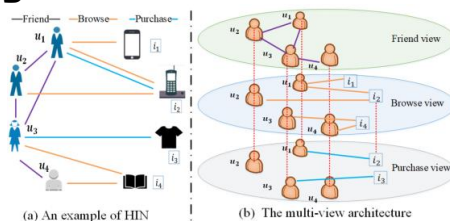
个人轨迹、行程、POI签到的私有性 (差分隐私)



时空数据的对抗攻击性

人脸识别的有效性与安全性, 对抗造假的能力
(深度学习鲁棒性)

异质图信息网络、元学习 (Meta learning)





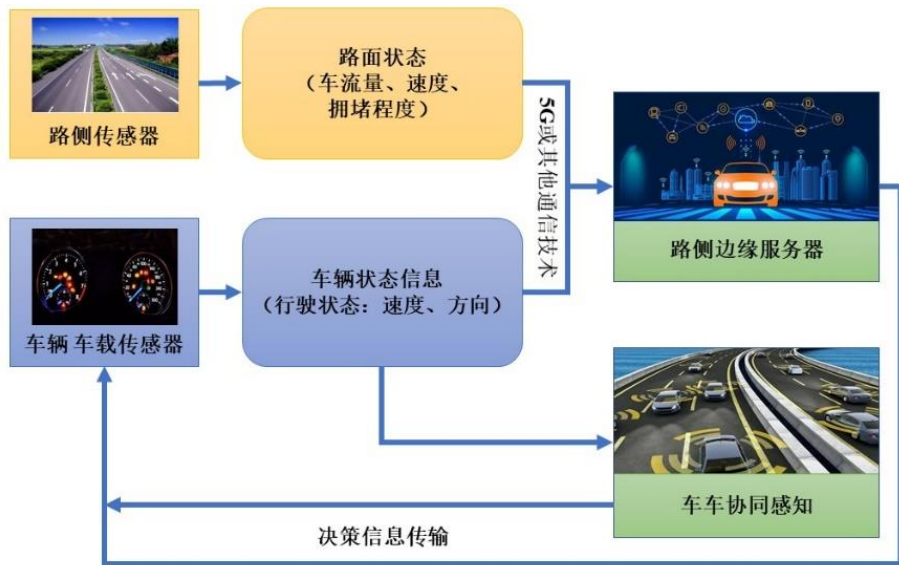
时空数据挖掘的前沿研究

云计算与边缘计算

减轻终端计算负载，将部分共性的识别、发现和决策的计算负载迁移至边缘服务器，由其完成智能辅助驾驶的相关计算。

典型应用：车路协同

- 本地计算
- 任务调度
- 协同计算
- 目标：通信时间总和最短



时空数据挖掘的前沿研究

□ 时空数据的相关性与不确定性

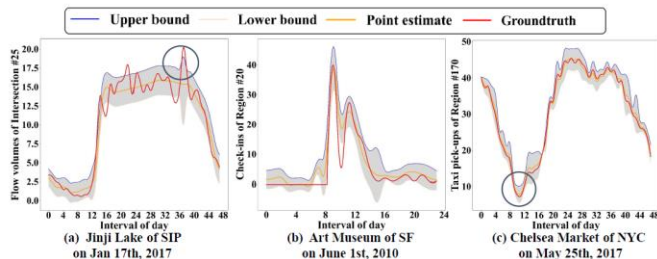
人类活动的本质规律 交通事件、城市新地点探索 => 不确定

Predictability & Uncertainty

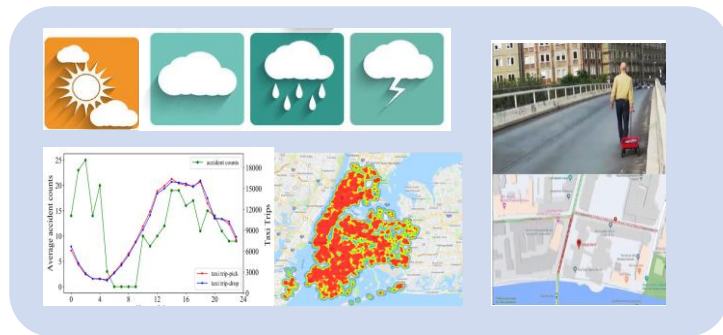
Intention & spatiotemporal context (Interactions)

认知不确定 Epistemic uncertainty (Model)

偶然不确定 Aleatoric uncertainty (Noise & data input)



□ 深度学习与时空数据挖掘的可解释性研究



心得体会-Take away message

新的研究，两个视角：

➤ 模型角度（一般到特殊）

- General模型的一般性缺点，改进模型，如GCN的可学习权值不再node-wise shared；增强原模型的表征能力；
- 根据改进技术再拓展到不同的应用层面（多样化）

➤ 问题角度（特殊到一般）

- 从问题出发，对于特定的任务，从数据中挖掘技术路线
- 再找类似的问题，对模型推广验证

心得体会-Take away message

深度学习应用研究核心要素

1. 数据是什么：输入、输出 – 数据预处理
2. 模型
 - 模型结构：数据流动与组合 (concat, LSTM aggregation)
 - 损失函数：使模型学习某种信息，进而可以捕获某种物理意义（语义信息）
3. 优化方法：Adam, SGD, AdaGrad等

心得体会-Take away message

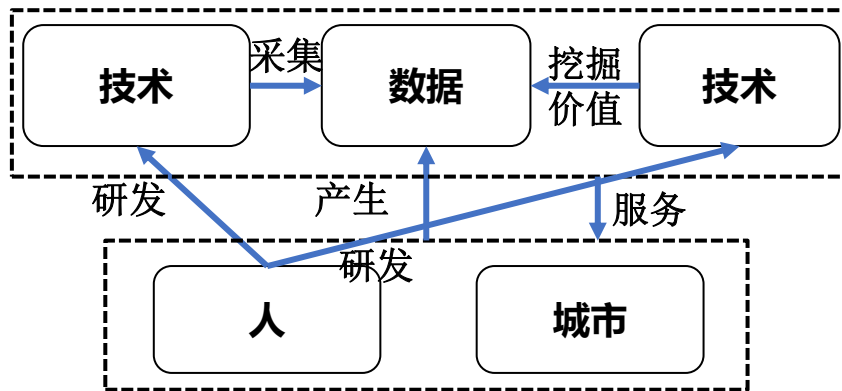
论文阅读与写作

- Decoder & Encoder
- 面到点，点到面 即钻进去再钻出来，循环上升
- 不局限自己研究的领域，**广泛阅读->抓共性问题，找相似**

心得体会-Take away message

技术，人与城市：

- 算法、数据服务于人类幸福与城市智慧化
- 技术推动城市治理精细化、科学化、现代化



经济价值

社会价值

事故预测
轨迹预测
订单匹配
房价预测

环境估计
流量推断
不确定性估计
可预测性分析



Selected Publication

【1】 **Zhou Z**, Wang Y, Xie X, et.al. “RiskOracle: A Minute-level Citywide Traffic Accident Forecasting Framework.” 2020 34th AAAI, 2020.

【2】 **Zhou Z**, Wang Y, Xie X, et.al. “Foresee Urban Sparse Traffic Accidents: A Spatiotemporal Multi-Granularity Perspective.” IEEE Transactions on Knowledge and Data Engineering (TKDE), 2021.

【3】 Wang Y, **Zhou Z**, Liu K, et.al. “Large-Scale Intelligent Taxicab Scheduling: A Distributed and Future-Aware Approach.” IEEE Transactions on Vehicular Technology (TVT), 2020.

【4】 Shao K, Wang Y, **Zhou Z**, Xie X, et.al. “TrajForesee: How limited detailed trajectories enhance large-scale sparse information to predict vehicle trajectories?” 2021 37th IEEE ICDE, 2021, Chania, Greece.

THANKS!

Q&A

祝大家学有所成，Paper多多，收获满满！

中科大计算机学院 周正阳

学术主页：

<http://staff.ustc.edu.cn/~angyan/pub.html> 汪炆

<http://home.ustc.edu.cn/~zzy0929/Home/> 周正阳