京东城市
JD iCity

Urban Computing

Building Intelligent Cities with Big Data and AI

Prof. Dr. Yu Zheng

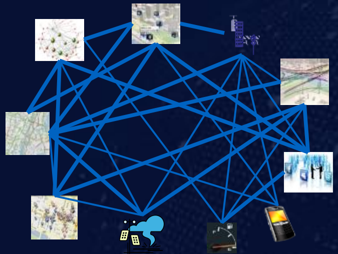
Vice President of JD.COM， Director of JD Intelligent City Research
Editor-in-Chief of ACM Transactions on Intelligent Systems and Technology (TIST)

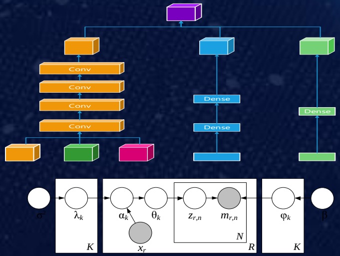
http://icity.jd.com

Urban Computing

京东城市
JD iCity

Zheng, Y., et al. Urban Computing: concepts, methodologies, and applications. *ACM TIST*.





Cloud Computing + **Big Data** + **AI**

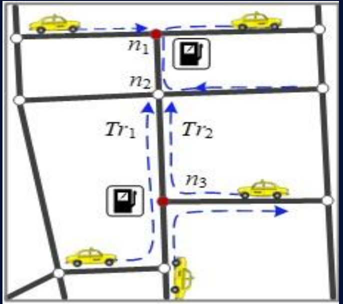
+

Cities

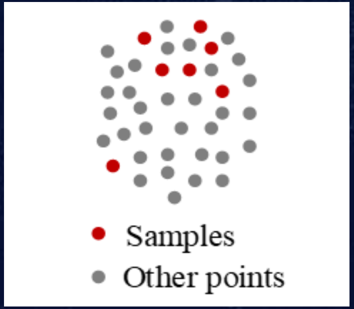
Urban Sensing

Urban Sensing

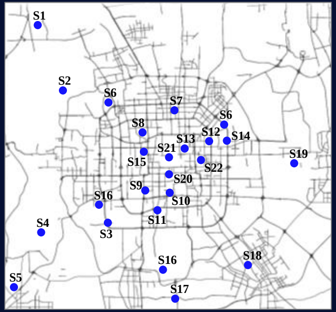
- Challenges & Research Topics



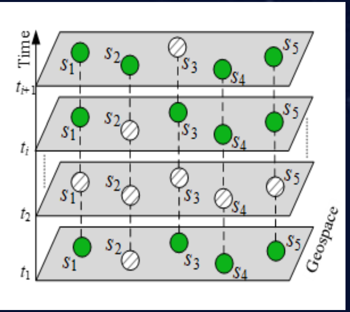
Resource deployment



Sampled data → Biased distribution



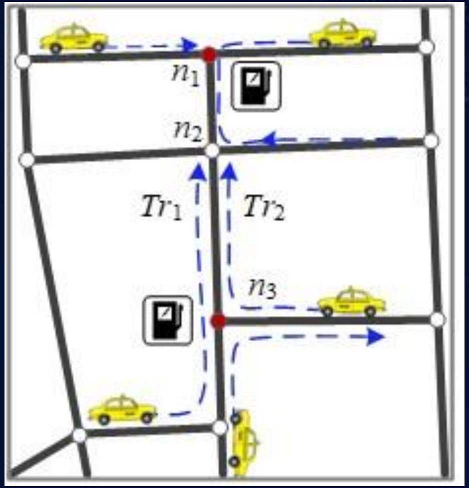
Data sparsity



Data missing

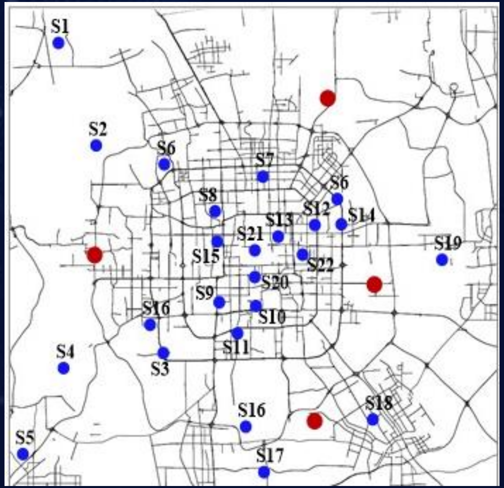
Urban Sensing Challenges- Resource deployment

Candidate selection is could be a NP-hard Problem



Y. Li, et al. Mining the Most Influential k-Location Set From Massive Trajectories. IEEE Transactions on Big Data. 2017

Hard to define a measurement for evaluating the deployment

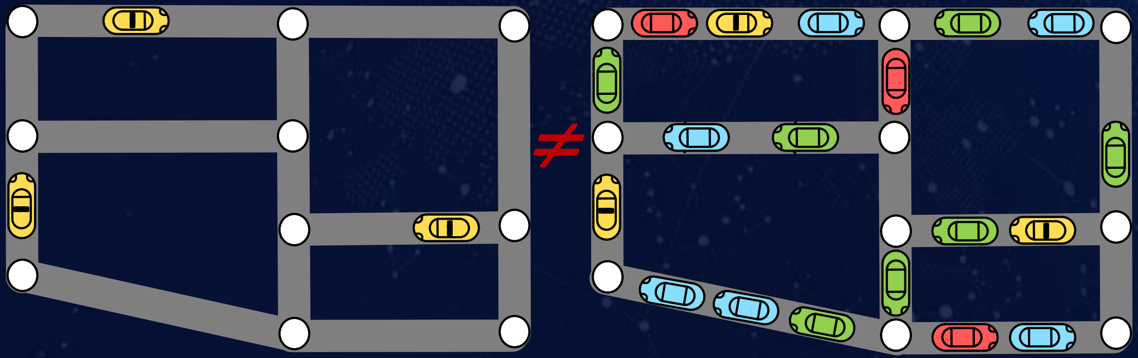


Hsieh H.P, S.D. Lin & Yu Zheng [KDD2015]

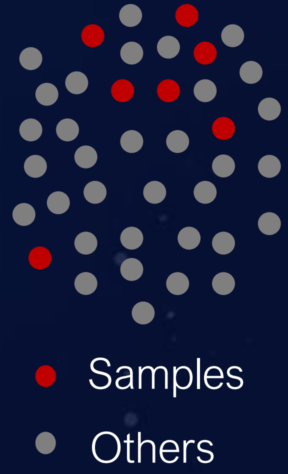
京东城市 JD iCity

Urban Sensing Challenges- Biased Samples

Sampled Data → Biased Distribution



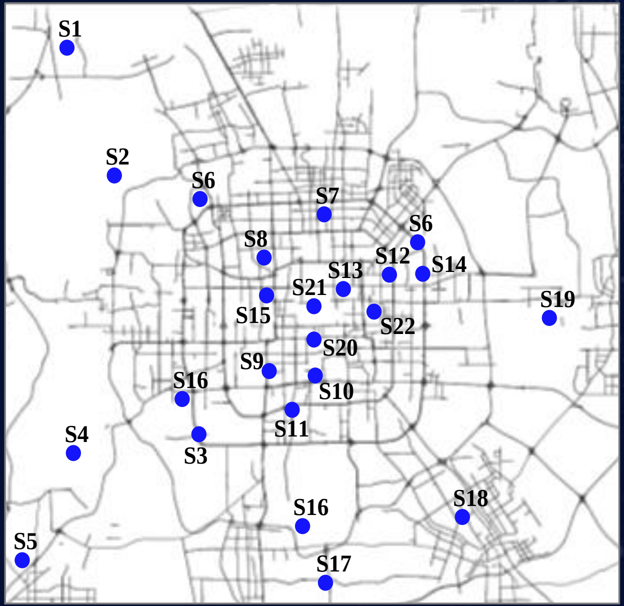
Taxi Flow → Entire Traffic Flow



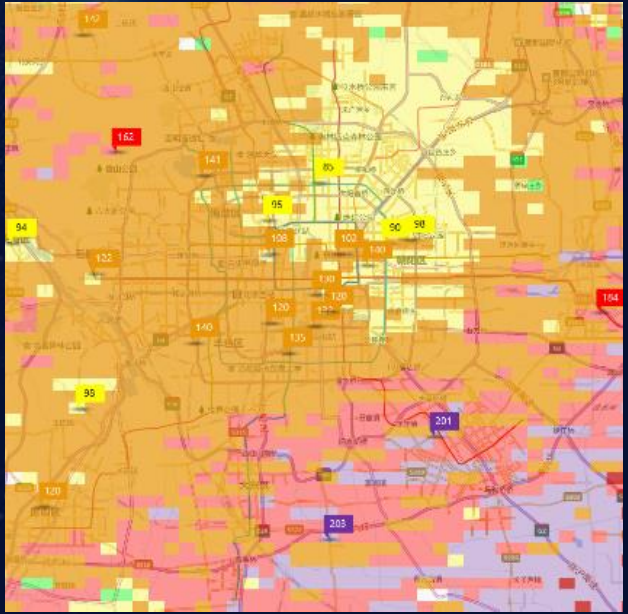
Jingbo Shang, Yu Zheng, et al. Inferring Gas Consumption and Pollution Emission of Vehicles throughout a City, KDD 2014

Urban Sensing Challenges - Data Sparsity

京东城市
JD iCity



Limited Air Quality Sensors



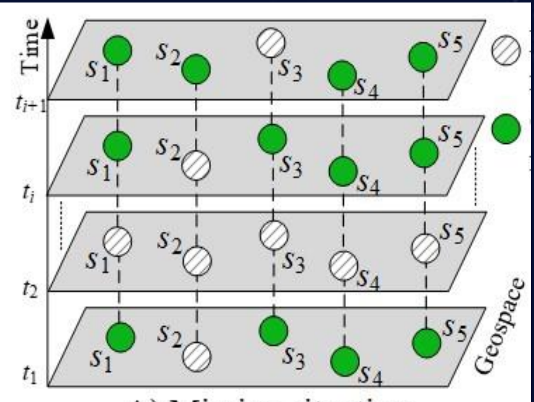
Fine-grained Air Quality throughout a City

Zheng, Y., et al. U-Air: When Urban Air Quality Inference Meets Big Data. KDD 2013

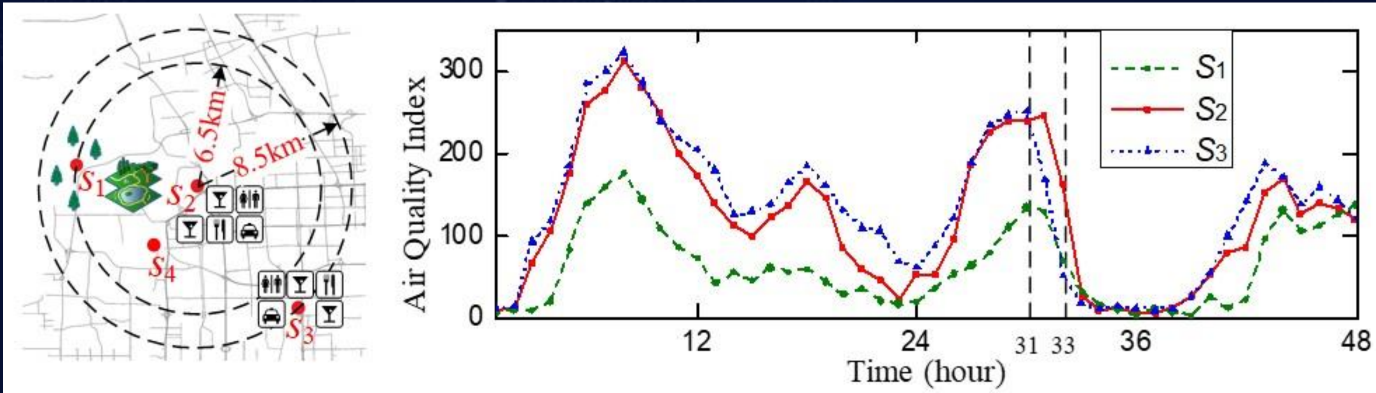
Urban Sensing Challenges - Data Missing

京东城市
JD iCity

- Lost data that is supposed to receive because of communication or device errors
- Difficult to fill
  - Randomly missing and block missing
  - Readings changing over time and location non-linearly



A) Missing situation



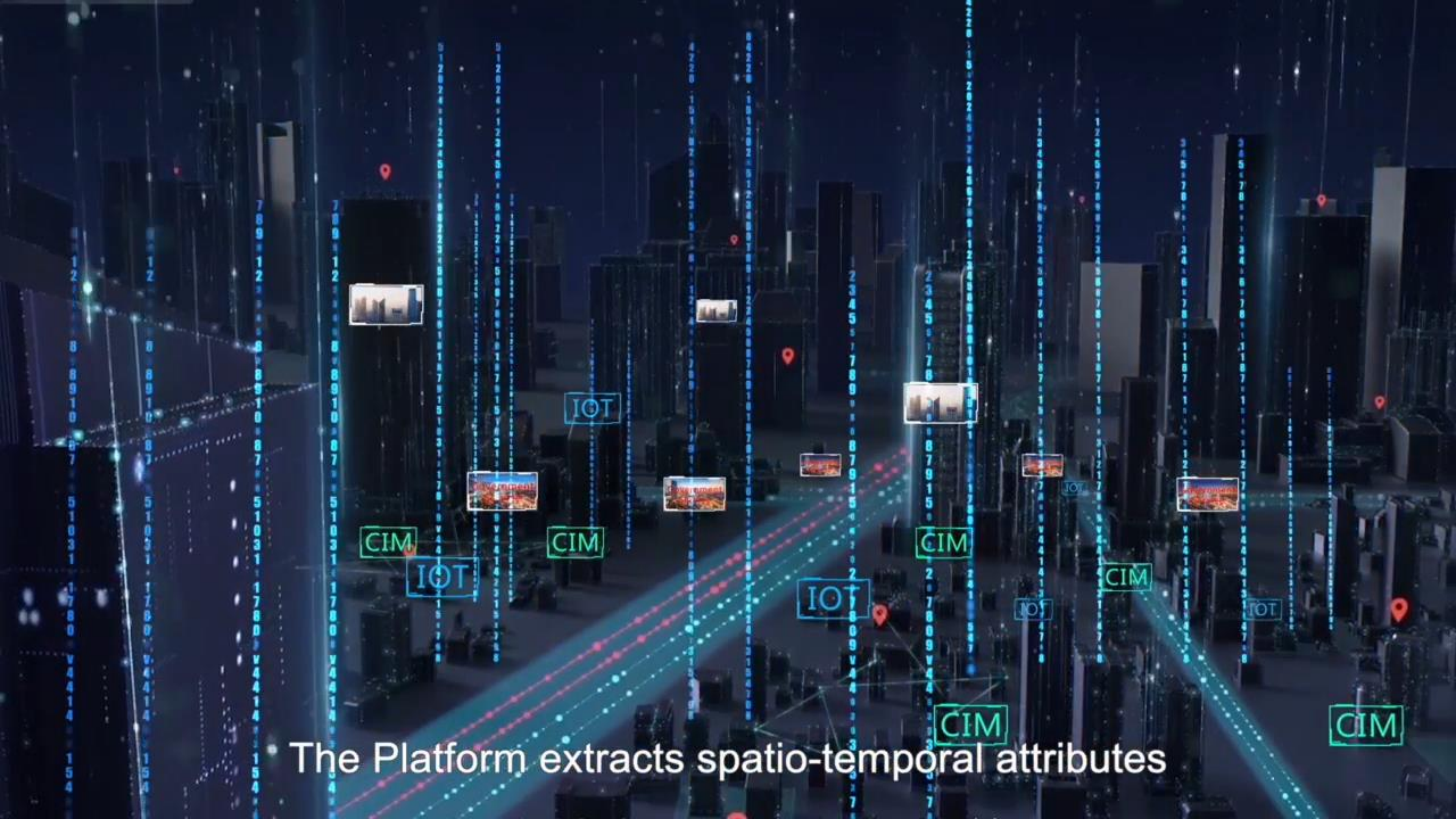
A) Geo-location of sensors B) Air quality index over time

Xiuwen Yi, **Yu Zheng**, et al. ST-MVL: Filling Missing Values in Geo-sensory Time Series Data. IJCAI 2016

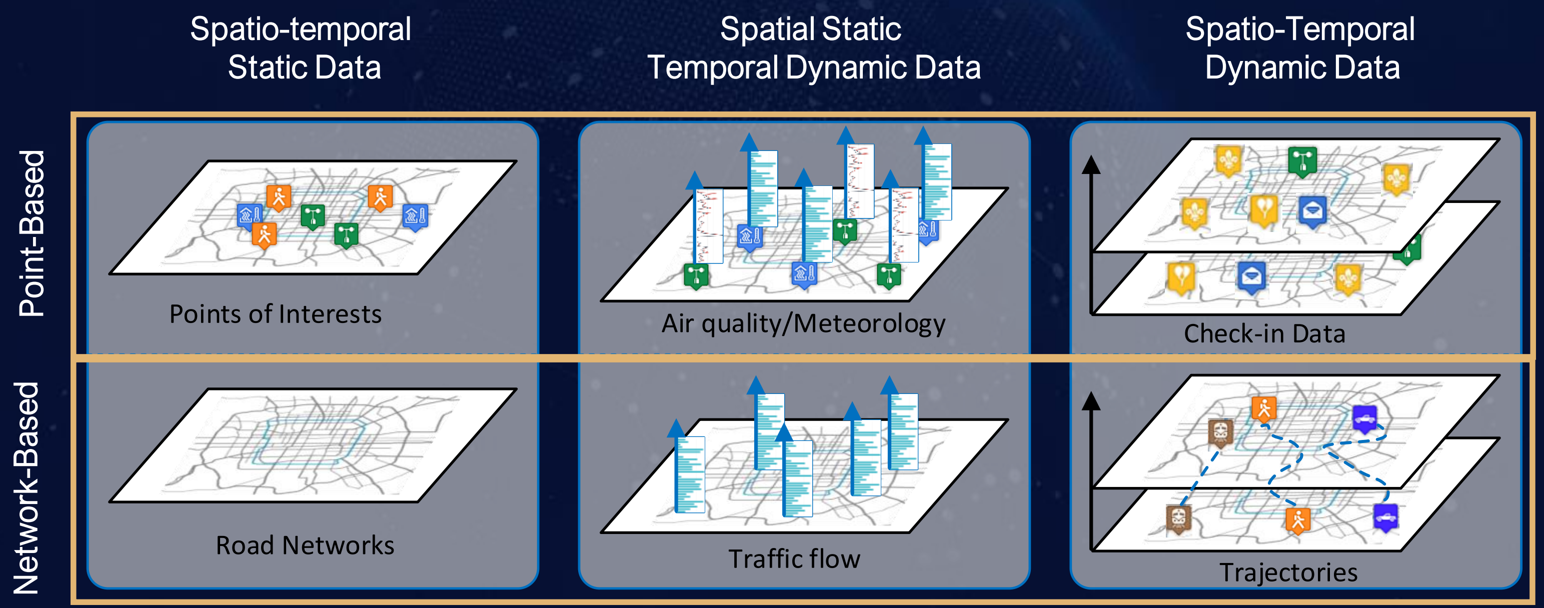
Urban Big Data

- Urban Data: Why unique
  - Data structures
  - Spatio-Temporal dynamics

Spatio-temporal Data


IOT
CIM
The Platform extracts spatio-temporal attributes

Spatiotemporal Data in Cities

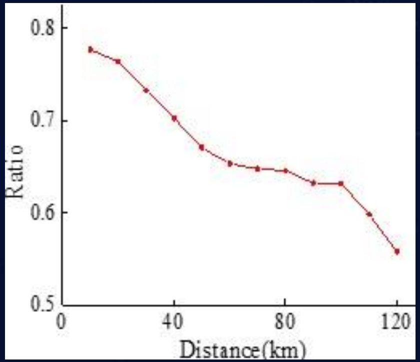


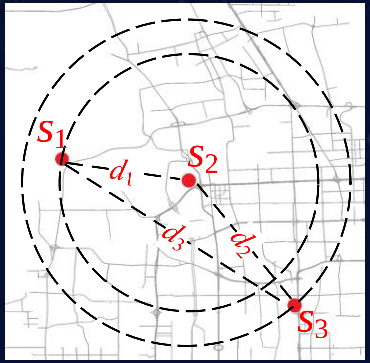
Yu Zheng. Trajectory Data Mining: An Overview. ACM Transactions on Intelligent Systems and Technology. 2015, vol. 6, issue 3.

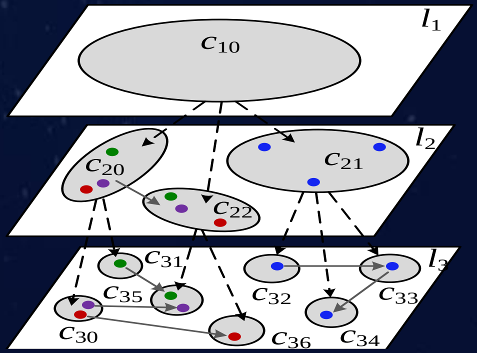
Why Spatiotemporal Data is unique

- Spatial Properties

- Spatial Distance
  - Spatial correlation
  - Triangle inequality: $|d_1 - d_2| \le d_3 \le |d_1 + d_2|$
- Spatial hierarchy
  - Different spatial granularities
  - City structures





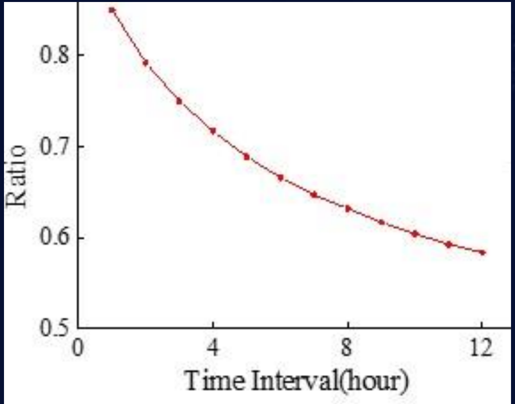


Junbo Zhang, **Yu Zheng**, et al. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction, AAAI 2017

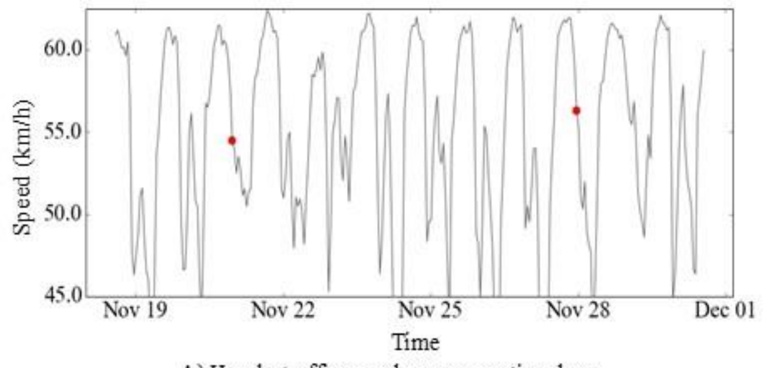
Why Spatiotemporal Data is unique

Temporal Properties

Temporal closeness

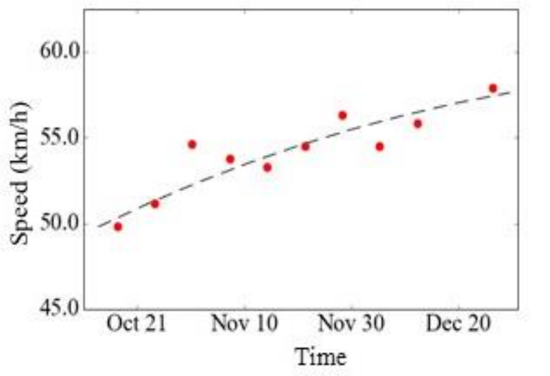


Period



A) Hourly traffic speed on consecutive days

Trend



B) Traffic speed at 9-10am on consecutive weekends

Junbo Zhang, **Yu Zheng**, et al. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction, AAAI 2017

Urban Data Mining

- Challenges to Data Management

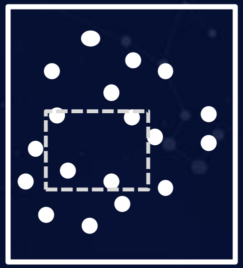
- Challenges to Data Analytics

Challenges Posed to Urban Data Management

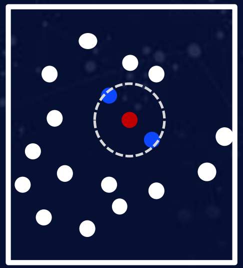
- Large-scale and highly dynamic → cloud computing
- Cloud computing platforms do not support ST Data well
  - Unique ST data structures: trajectories (the most complex ST Data)
  - Unique queries: ST-Range queries and KNN queries rather than key words
  - Data across different domains: Hybrid indexing for managing multi-modality data

Trajectories

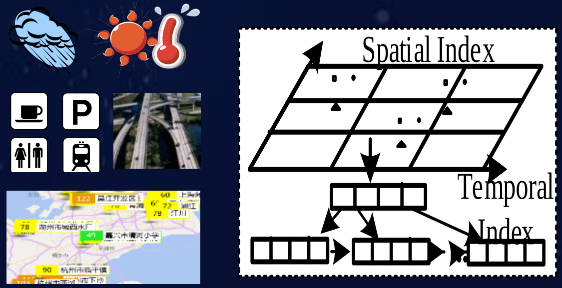
Range Queries



KNN Queries



Hybrid indexing



Jie Bao, et al. Managing Massive Trajectories on the Cloud, ACM SIGSPATIAL 2016

Example 1

Detecting Vehicle Illegal Parking Using Sharing Bikes' Trajectories

Detecting Vehicle Illegal Parking Using Sharing Bikes' Trajectories

T. He, J. Bao, R. Li, S. R., Y. Li, C. Tian, Y. Zheng. Detecting Vehicle Illegal Parking Events using Sharing Bikes' Trajectories. KDD 2018

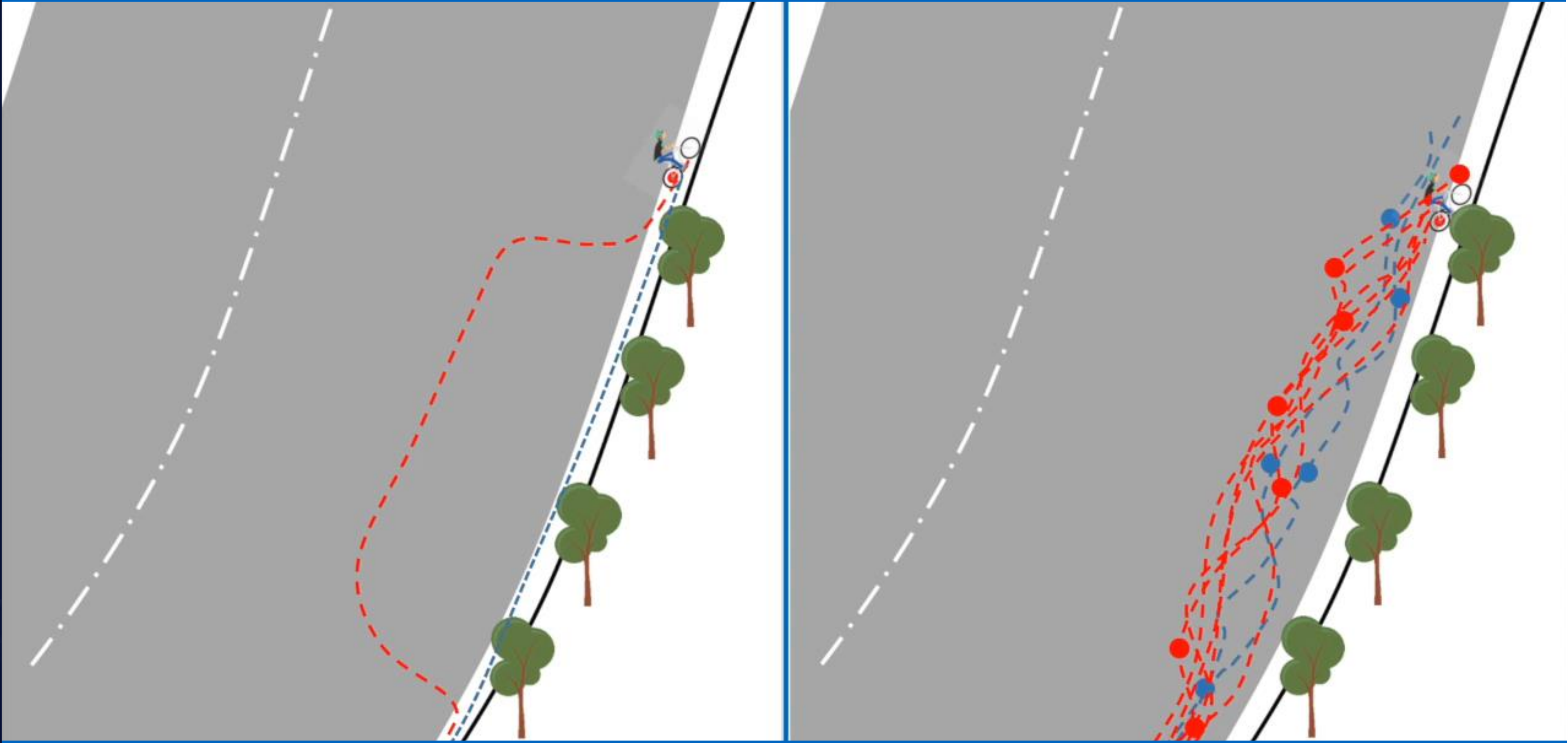
Detecting Vehicle Illegal Parking Using Sharing Bikes’ Trajectories

T. He, J. Bao, R. Li, S. R., Y. Li, C. Tian, Y. Zheng. Detecting Vehicle Illegal Parking Events using Sharing Bikes' Trajectories. KDD 2018

Detecting Vehicle Illegal Parking Using Sharing Bikes' Trajectories

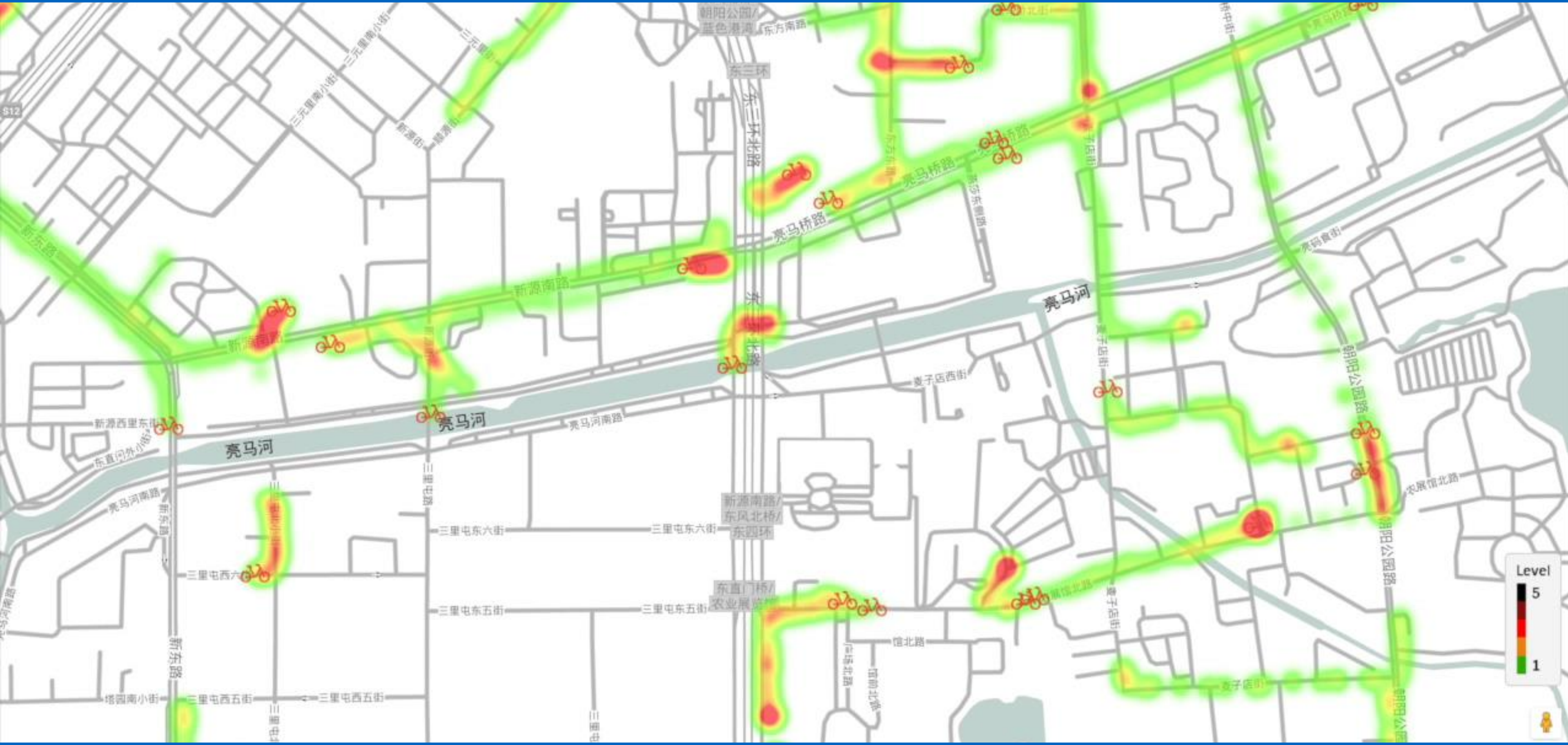
T. He, J. Bao, R. Li, S. R., Y. Li, C. Tian, Y. Zheng. Detecting Vehicle Illegal Parking Events using Sharing Bikes' Trajectories. KDD 2018

Detecting Vehicle Illegal Parking Using Sharing Bikes' Trajectories



T. He, J. Bao, R. Li, S. R., Y. Li, C. Tian, Y. Zheng. Detecting Vehicle Illegal Parking Events using Sharing Bikes' Trajectories. KDD 2018

Detecting Vehicle Illegal Parking Using Sharing Bikes' Trajectories



T. He, J. Bao, R. Li, S. R., Y. Li, C. Tian, Y. Zheng. Detecting Vehicle Illegal Parking Events using Sharing Bikes' Trajectories. KDD 2018

Challenges Posed to Urban Data Analytics

- Urban Data Analytics
  - Texts and images → spatial and spatiotemporal data; (encoding spatiotemporal properties)
  - Mining a single data source → Mining data across different domains
  - Pure machine learning → Visual and interactive data analytics

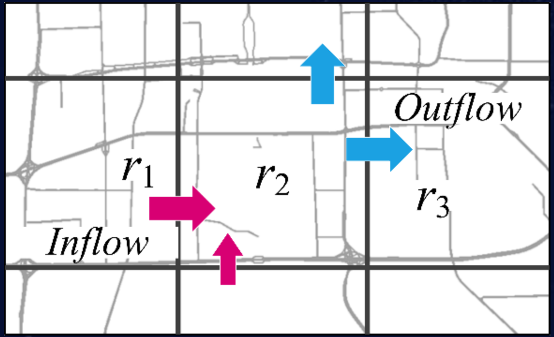
Challenge 1: Encoding spatiotemporal properties in machine learning models

Example

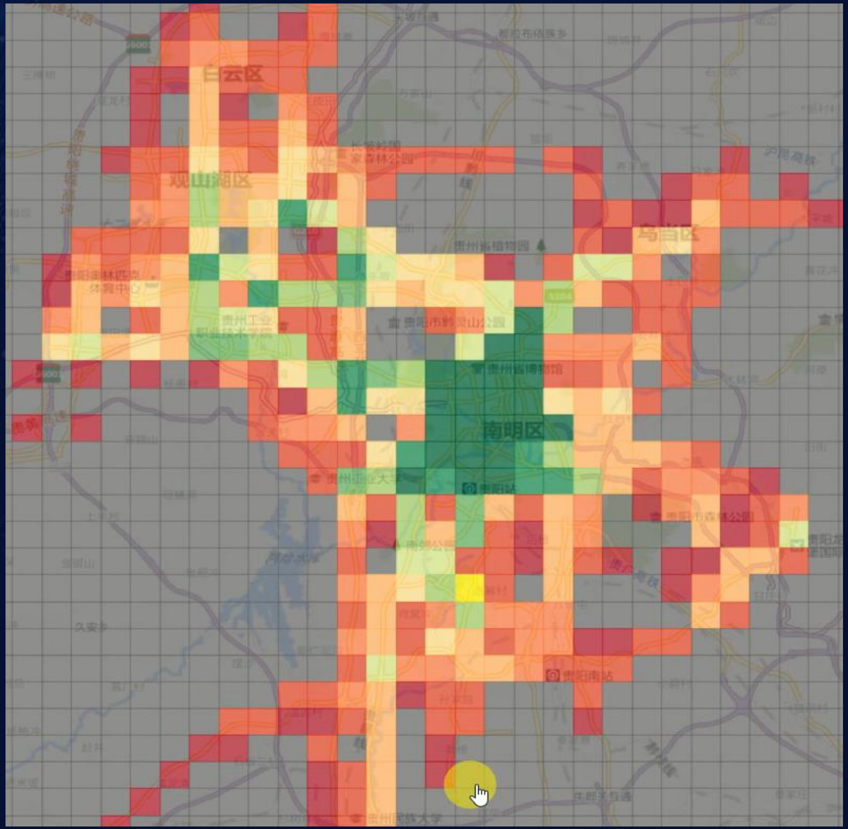
Crowd Flow Prediction

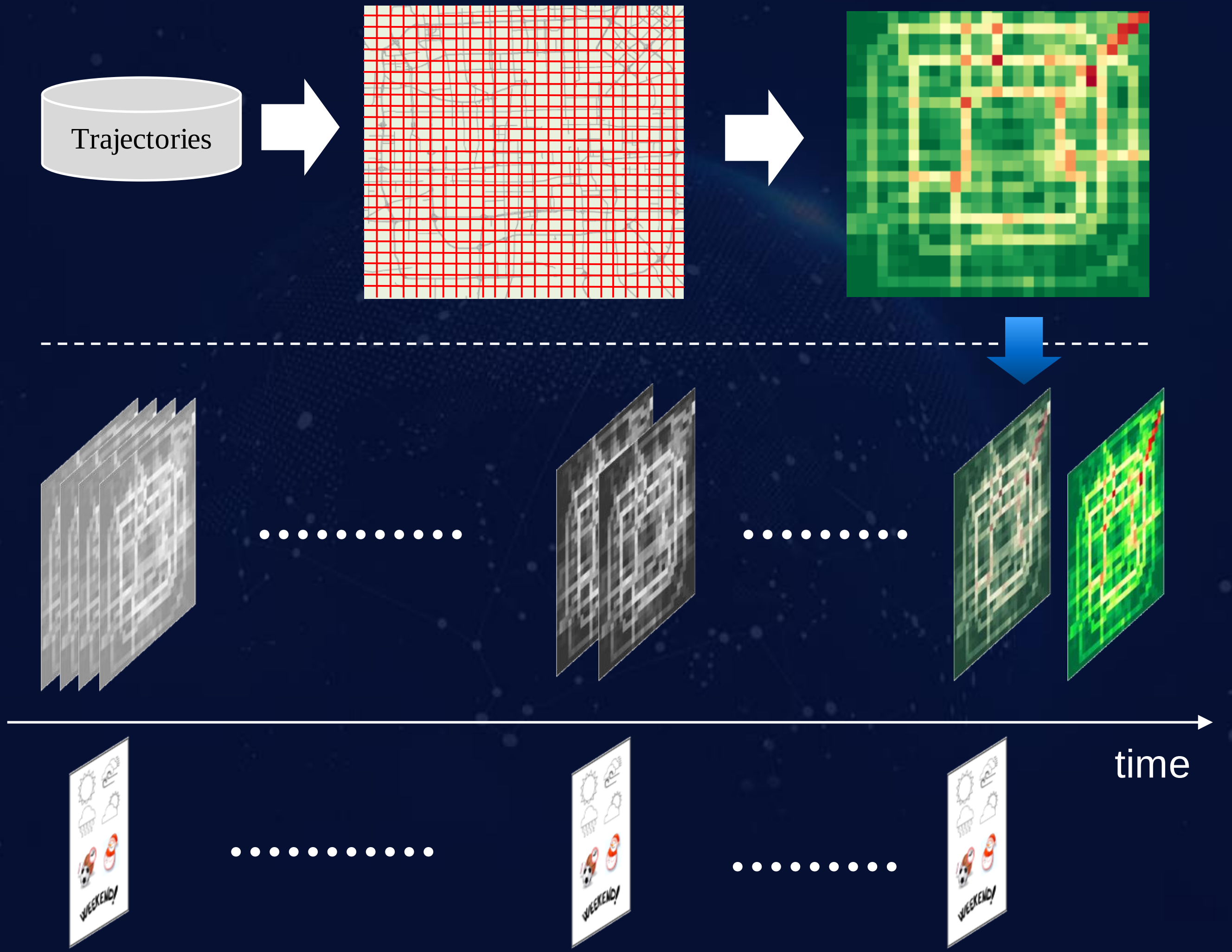
Predicting Crowd Flows throughout a City

- Predict in and out flow of crowds
  - In each grid
  - Over the next few hours



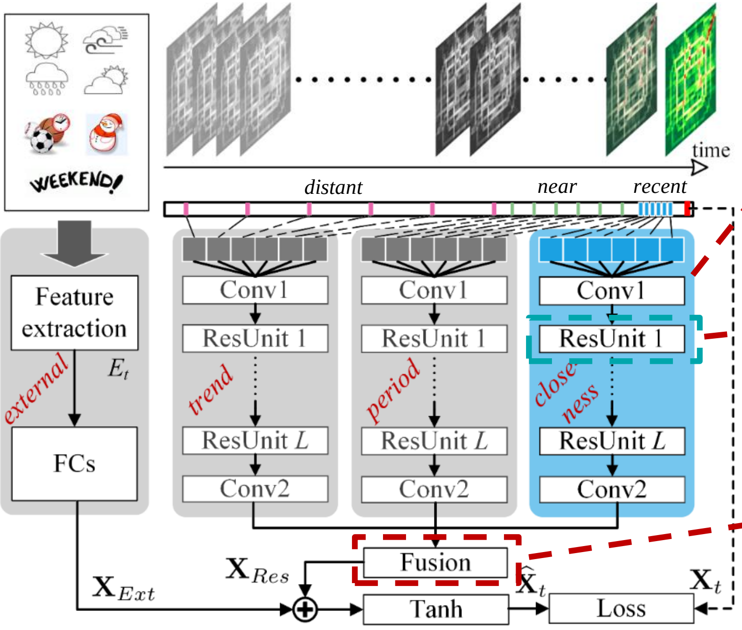
- Challenges
  - Complex and dynamic factors
  - Encoding spatiotemporal properties

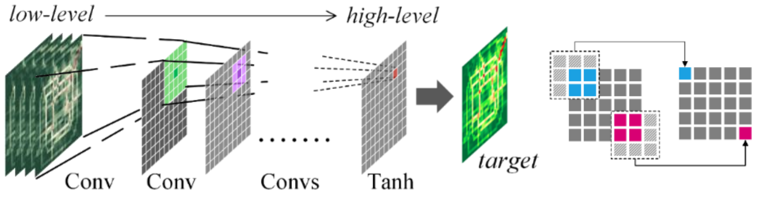



Trajectories
time
WEEKEND!
WEEKEND!
WEEKEND!

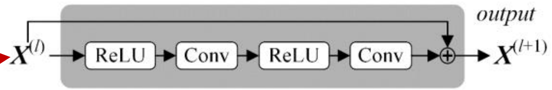
ST-ResNet Architecture: A Collective Prediction

- Capture temporal closeness, period, and trend
- Capture external factors
- Capture spatial correlation of both near and far distances





- Residual learning to help training



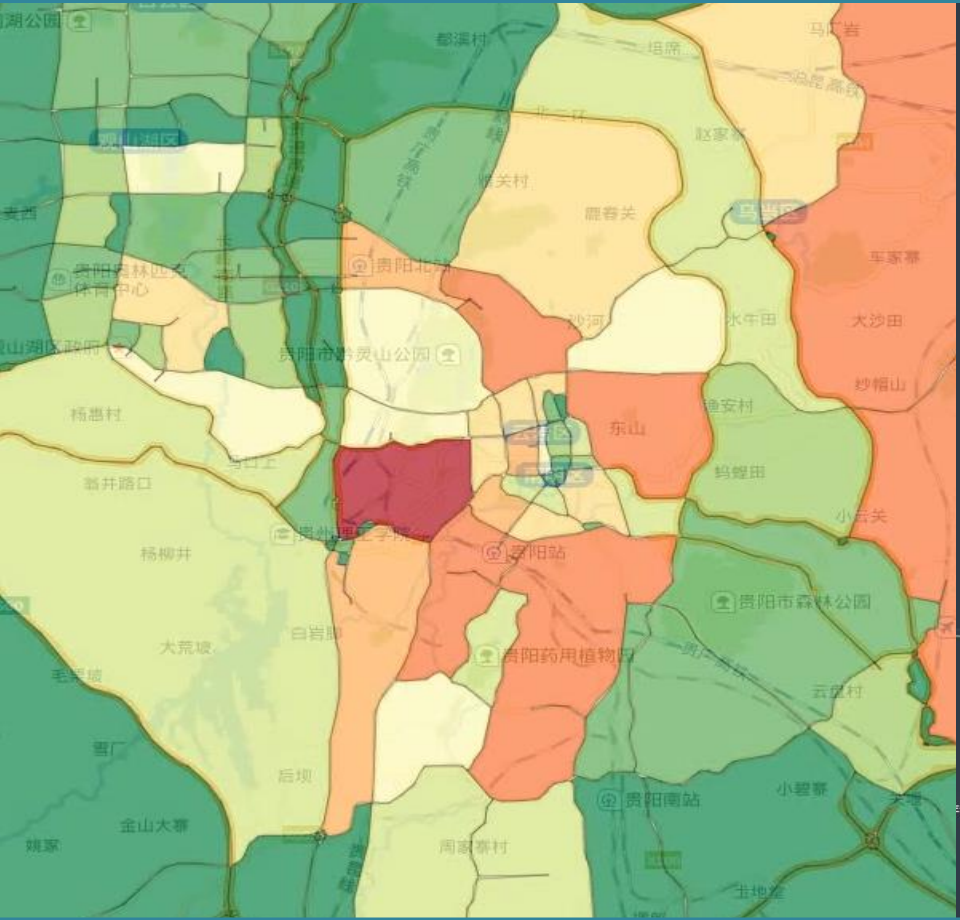
- Fusing factors differently in different regions

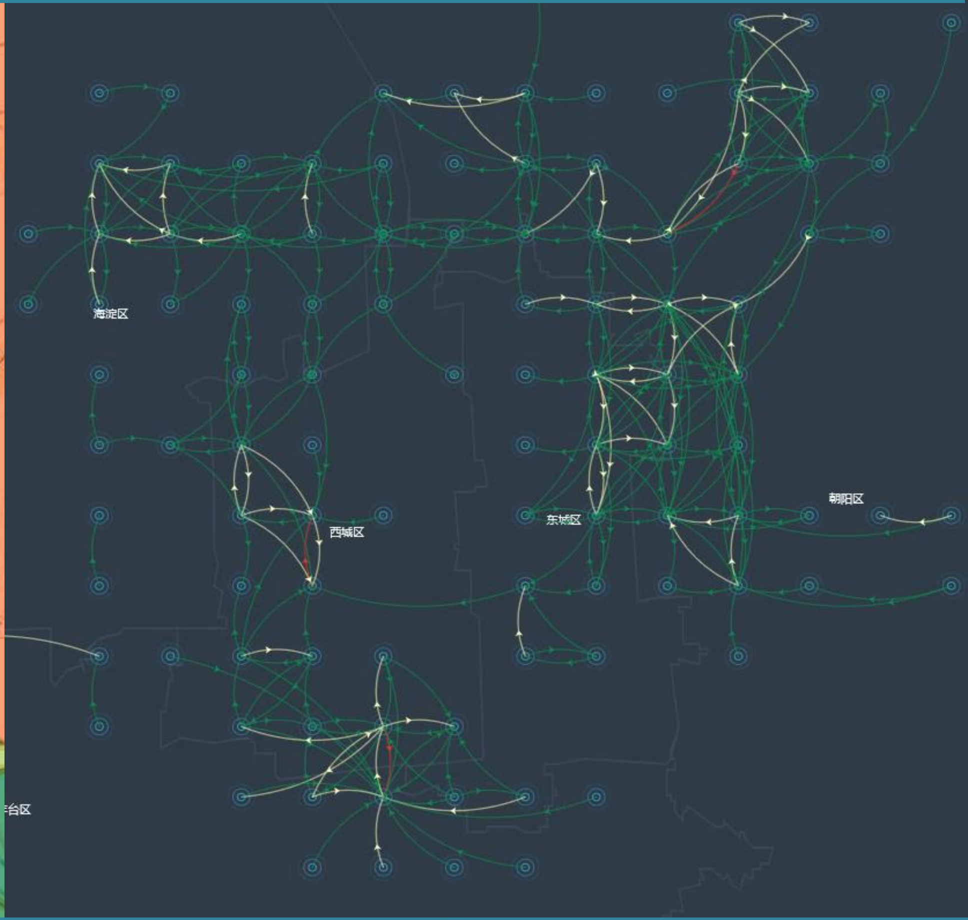
$$\begin{bmatrix} (\omega_{c,1},\omega_{p,1},\omega_{q,1}) & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & (\omega_{c,n},\omega_{p,n},\omega_{q,n}) \end{bmatrix}$$

$$\mathbf{X}_{Res} = \mathbf{W}_c \circ \mathbf{X}_c^{(L+2)} + \mathbf{W}_p \circ \mathbf{X}_p^{(L+2)} + \mathbf{W}_q \circ \mathbf{X}_q^{(L+2)}$$

Junbo Zhang, Yu Zheng, et al. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction, AAAI 2017

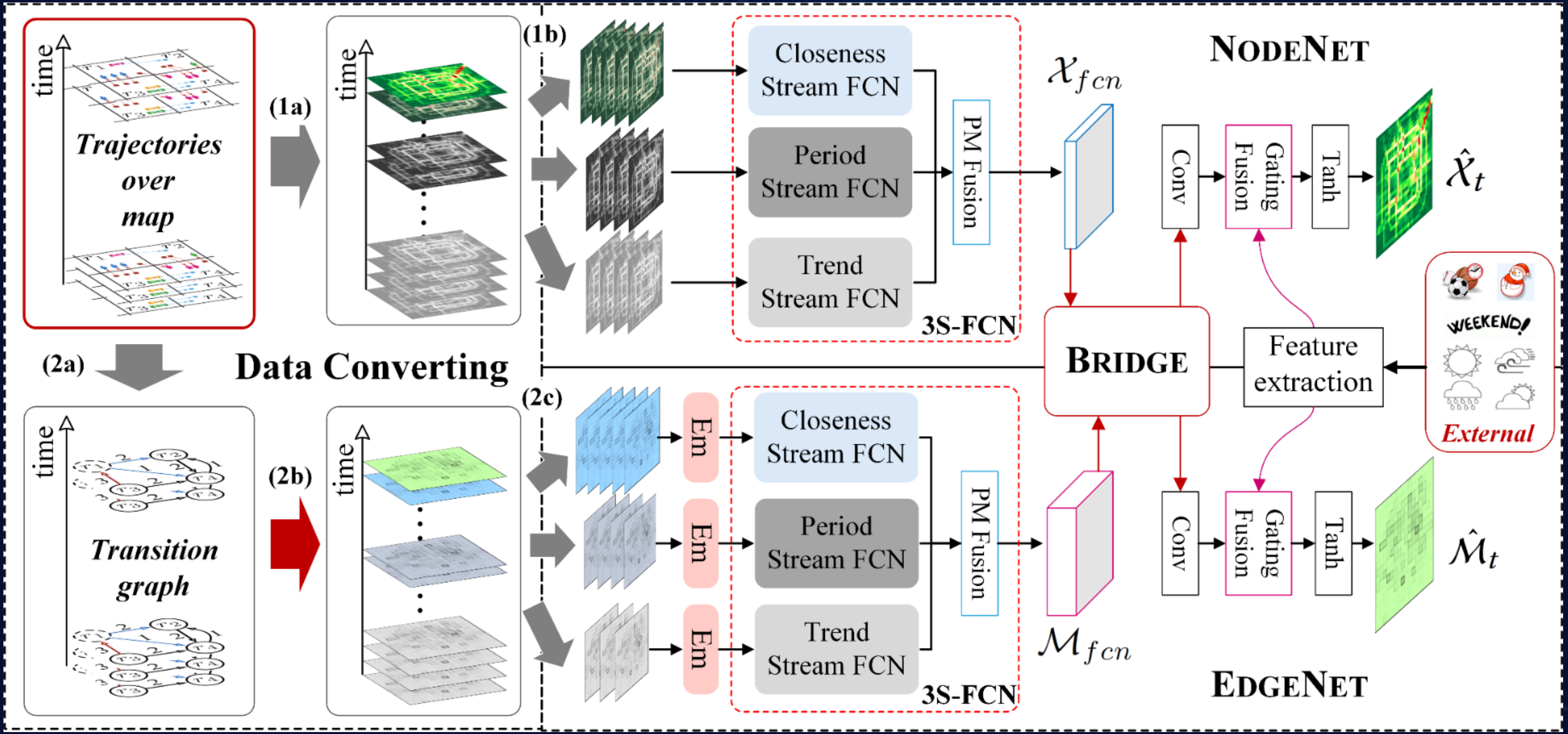
Predicting Crowd Flow Using AI





Junbo Zhang, Yu Zheng et al. Predicting citywide crowd flows using deep spatio-temporal residual networks. AI Journal

Multitask Deep Learning Framework



Junbo Zhang, Yu Zheng, et al. Flow Prediction in Spatio-Temporal Networks Based on Multitask Deep Learning, TKDE

Challenge 2: Mining data across different domains

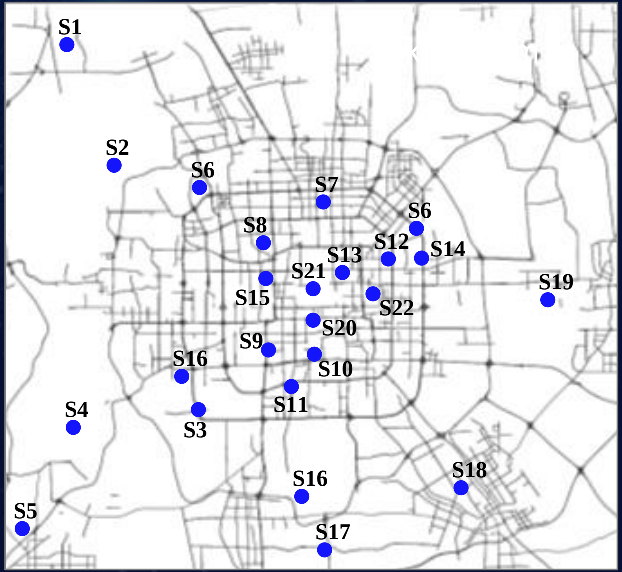
Example

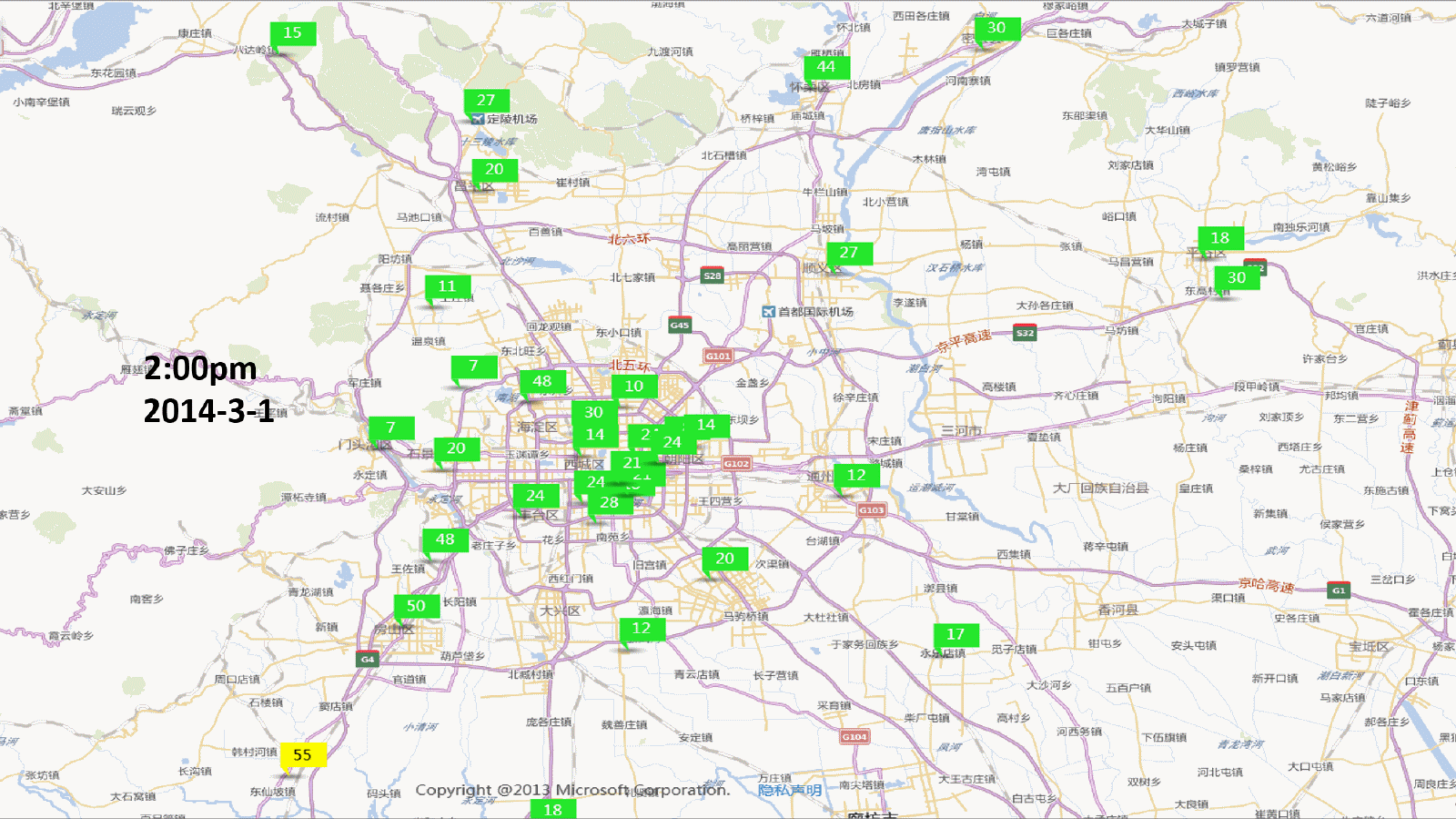
Air Quality Inference and Forecasting

Air Pollution: A Global Concern !

PM2.5, PM10, NO_2, SO_2, CO, O_3

- Air quality monitor station




2:00pm
2014-3-1
Copyright @2013 Microsoft Corporation.
隐私声明

Inferring **Real-Time** and **Fine-Grained** air quality throughout a city

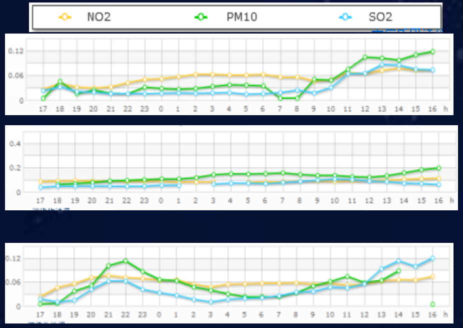
Meteorology

Traffic

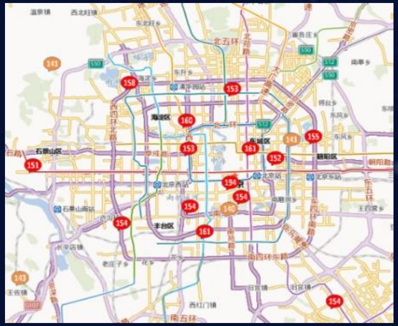
Human Mobility

POIs

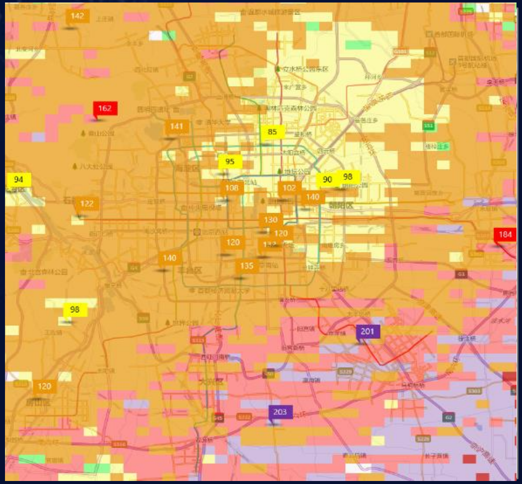
Road networks



Historical air quality data

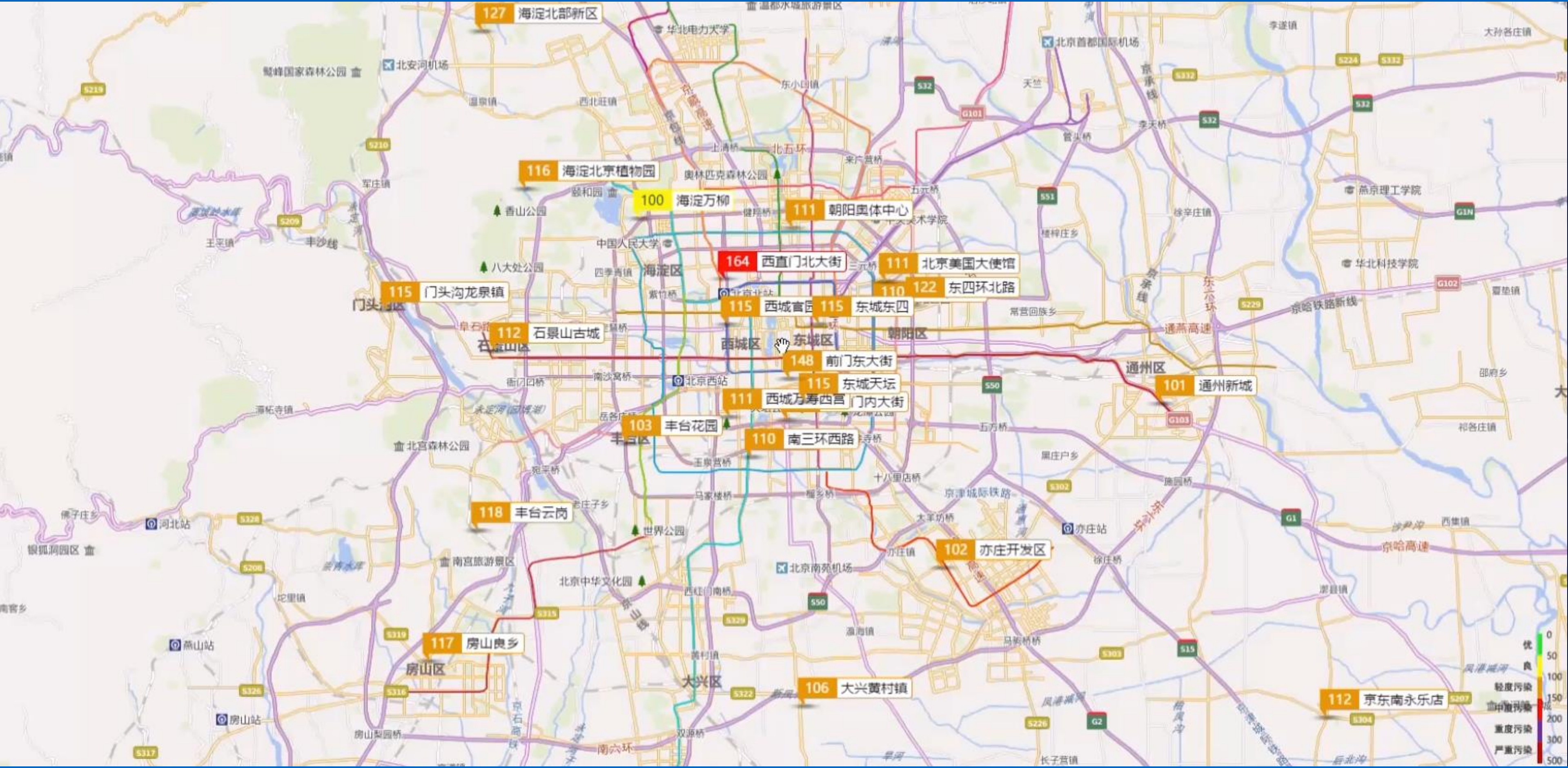


Real-time air quality reports



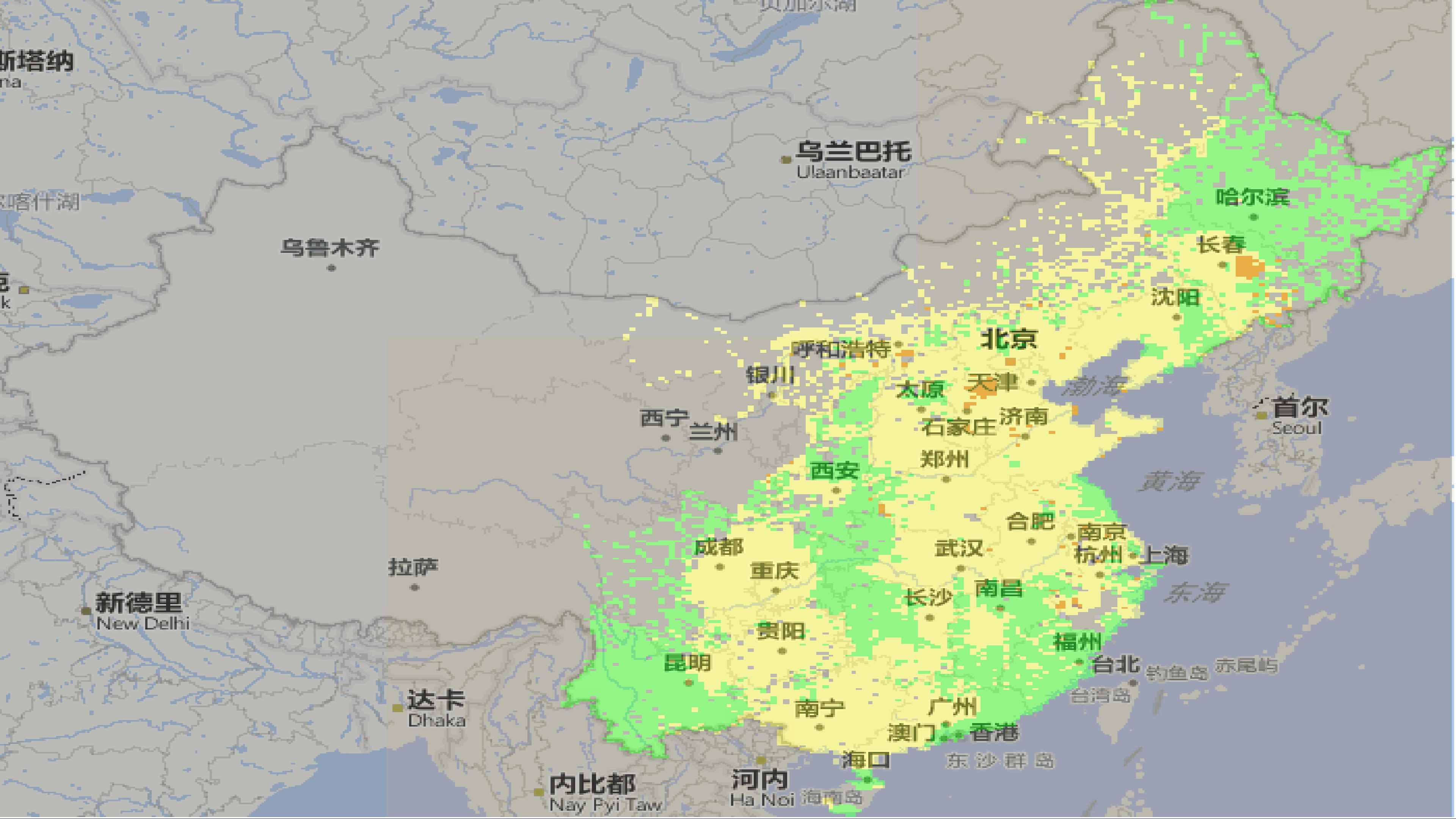
Zheng, Y., et al. U-Air: When Urban Air Quality Inference Meets Big Data. KDD 2013

Fine-Grained and Real-time Air Quality Inference



Zheng, Y., et al. U-Air: When Urban Air Quality Inference Meets Big Data. KDD 2013

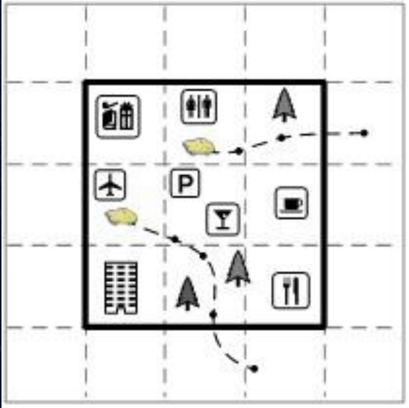

Intelligent Environment
and every local area despite incomplete air quality data

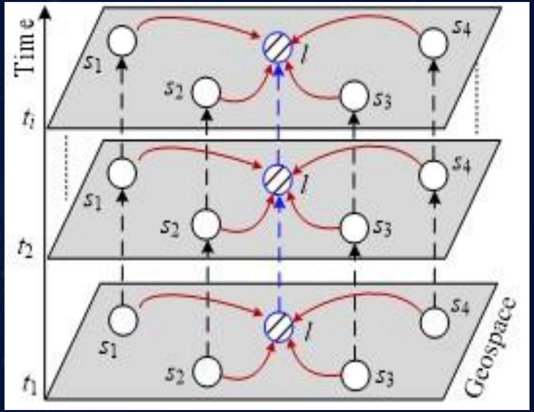


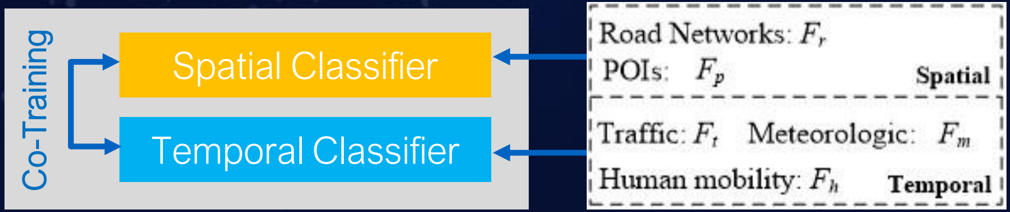
乌兰巴托
Ulaanbaatar
哈尔滨
长春
沈阳
乌鲁木齐
北京
呼和浩特
银川
太原
天津
渤海
首尔
Seoul
西宁
兰州
石家庄
济南
郑州
西安
黄海
合肥
南京
上海
杭州
武汉
成都
重庆
拉萨
东海
长沙
南昌
新德里
New Delhi
贵阳
福州
昆明
台北
钓鱼岛
赤尾屿
台湾岛
达卡
Dhaka
南宁
广州
澳门
香港
海口
东沙群岛
内比都
Nay Pyi Taw
河内
Ha Noi
海南岛

Semi-Supervised Learning Model

- Partition a city into disjoint grids
- Features from different datasets
- Encoding spatiotemporal properties
  - Temporal dependency in a location
  - Geo-correlation between locations
- Domain knowledge
  - Emission from a location
  - Propagation among locations
- Co-training-based inference model







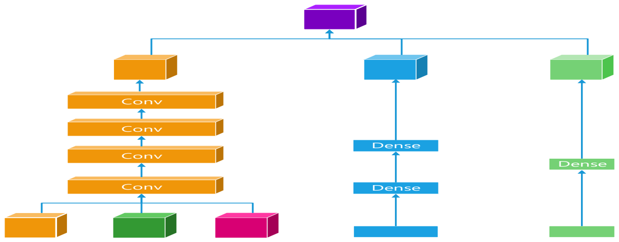
Zheng, Y., et al. U-Air: When Urban Air Quality Inference Meets Big Data. KDD 2013

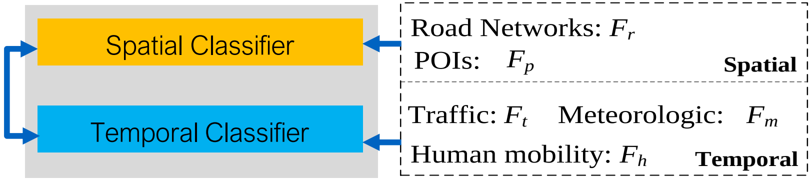
Forecast Air Quality over the next 48 hours

Xiuwen Yi, Junbo Zhang, et al. Deep Distributed Fusion Network for Air Quality Prediction. KDD 2018

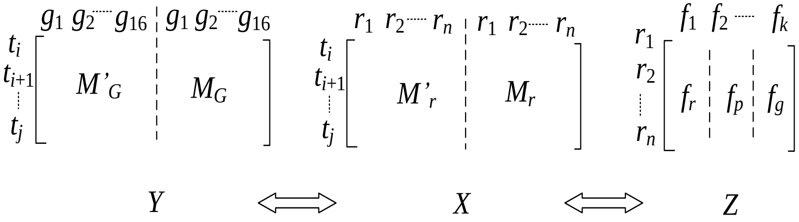
Methodologies for Cross-Domain Knowledge Fusion

- Stage-based data fusion
- Feature-level-based data fusion
  - Feature concatenation + regularization
  - DNN-based
- Semantic meaning-based fusion

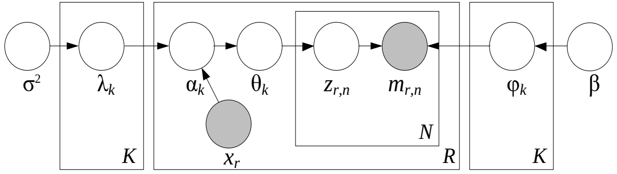




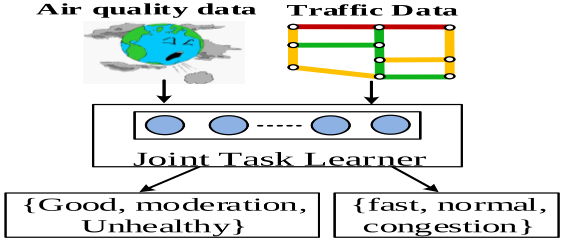
Multi-view learning (Co-training)



Similarity-based (matrix factorization)



Pro. dependency-based (Topic Models)



Transfer Learning-based

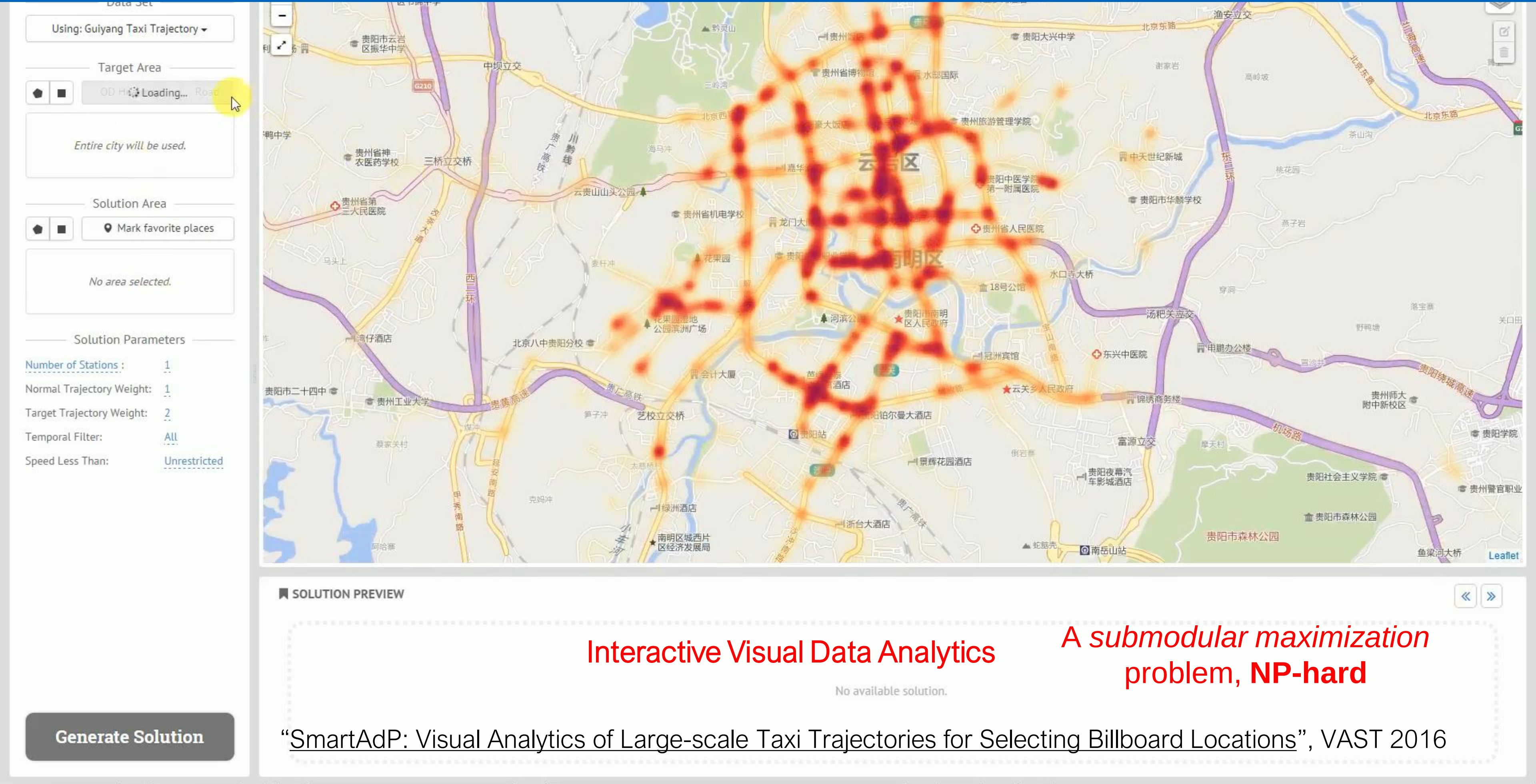
Yu Zheng. Methodologies for Cross-Domain Data Fusion: An Overview. IEEE Transactions on Big Data

Challenge 3: Visual and interactive data analytics

Example

Selecting locations for charging stations

Finding Top-k Most Influential Location Set



Interactive Visual Data Analytics

A *submodular maximization* problem, **NP-hard**

"SmartAdP: Visual Analytics of Large-scale Taxi Trajectories for Selecting Billboard Locations", VAST 2016

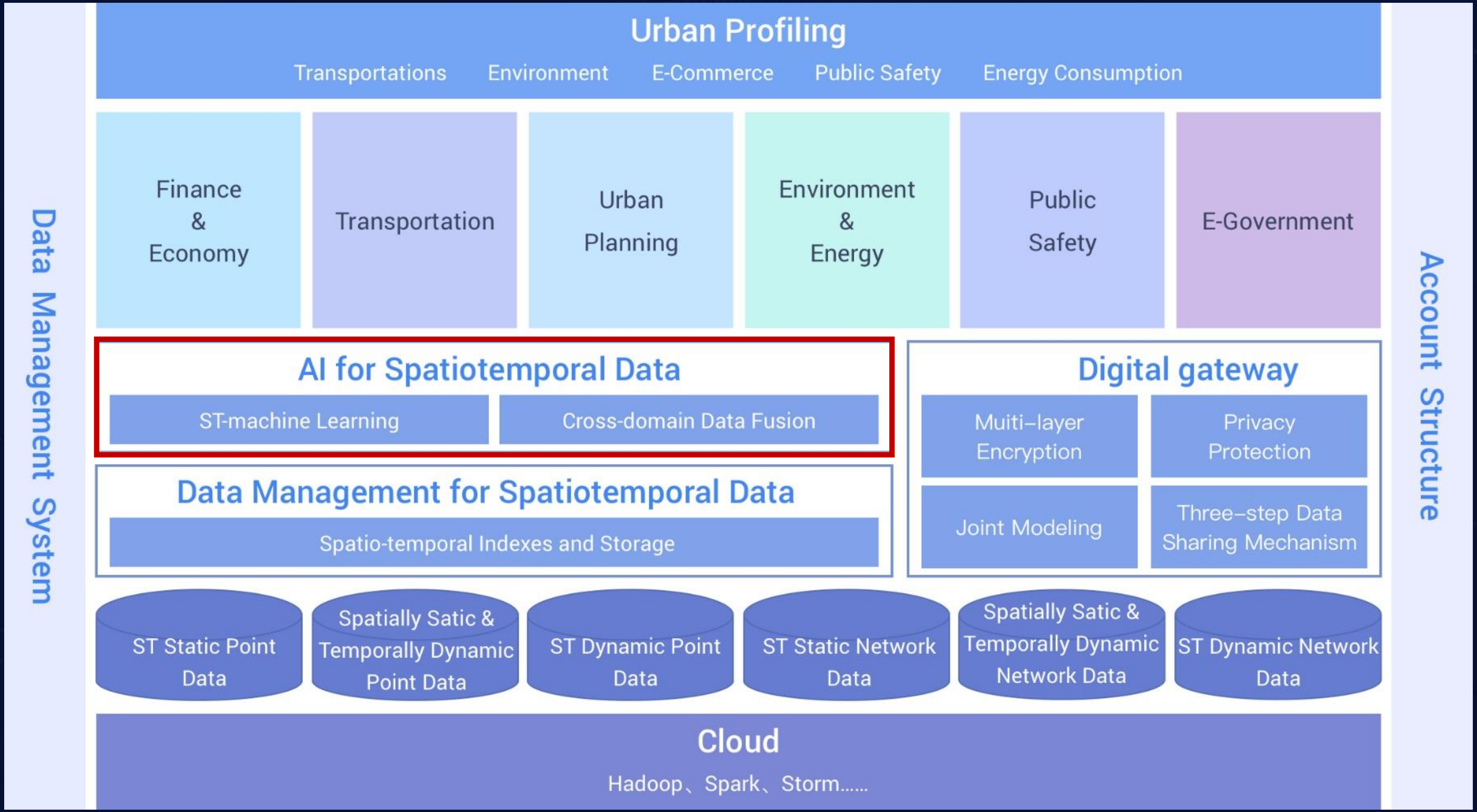
Providing Services

- Urban computing platforms
  - Six Spatio-temporal (ST) data models
  - ST data management
  - ST machine learning
  - Cross-domain machine learning
- Building City OS → an ecosystem

http://ucp.jd.com/

Urban Computing Platform

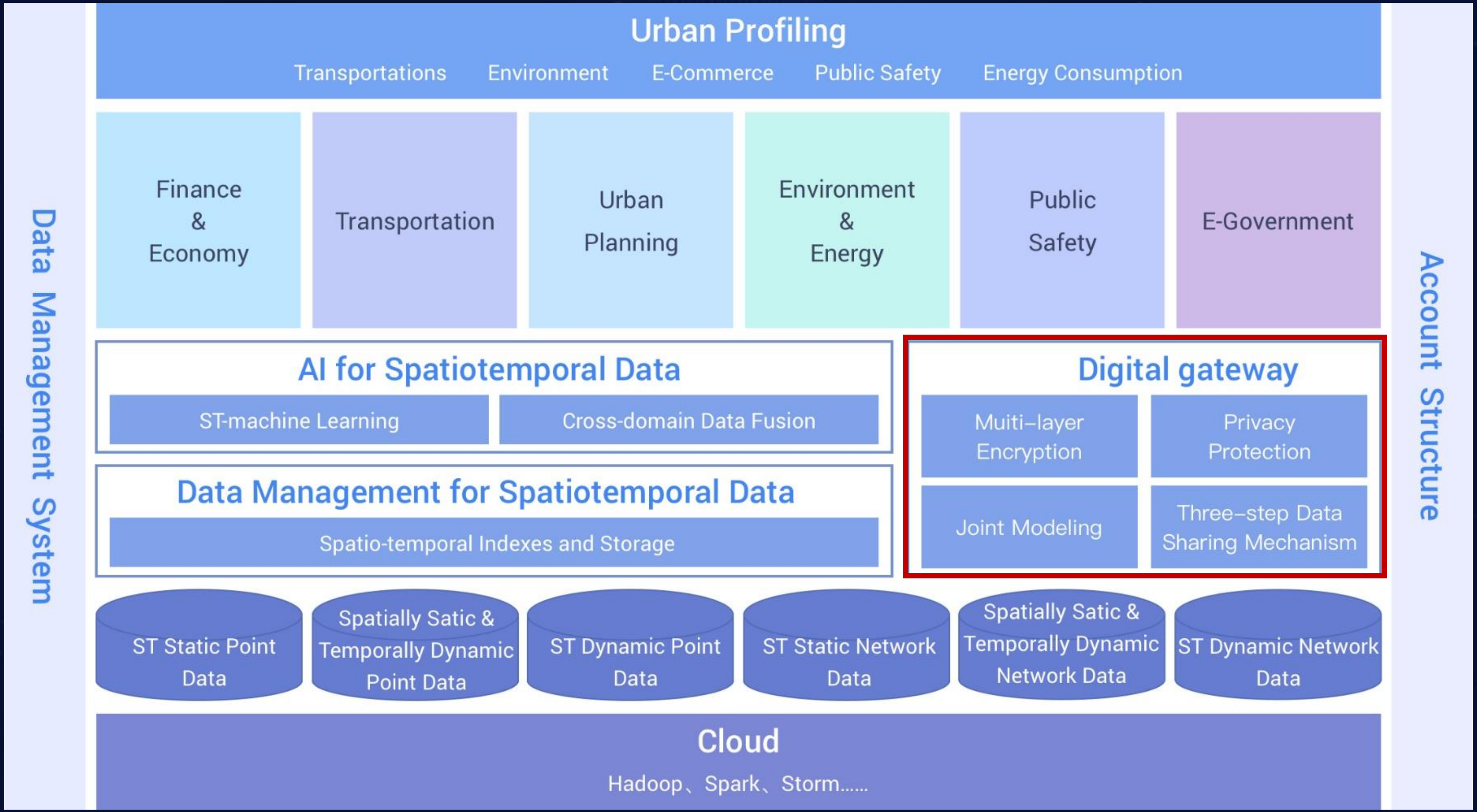
http://ucp.jd.com



http://ucp.jd.com

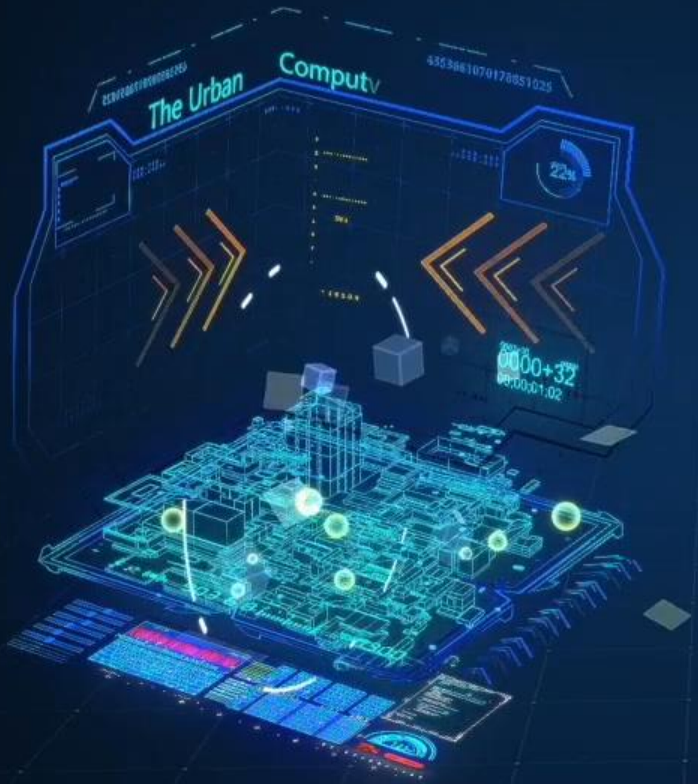
Urban Computing Platform

http://ucp.jd.com

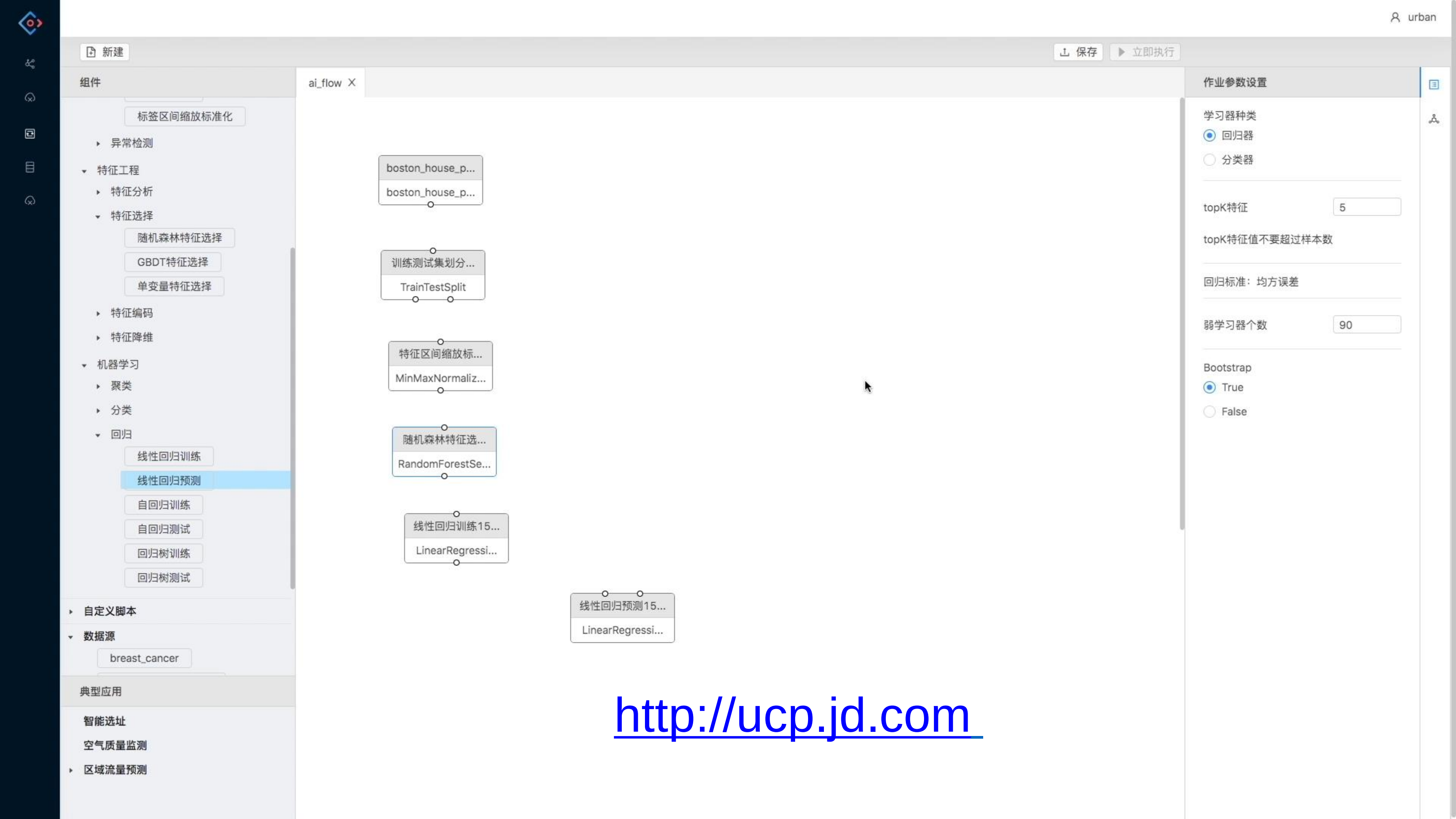



http://ucp.jd.com
The Digital Gateway on this platform ensures data security

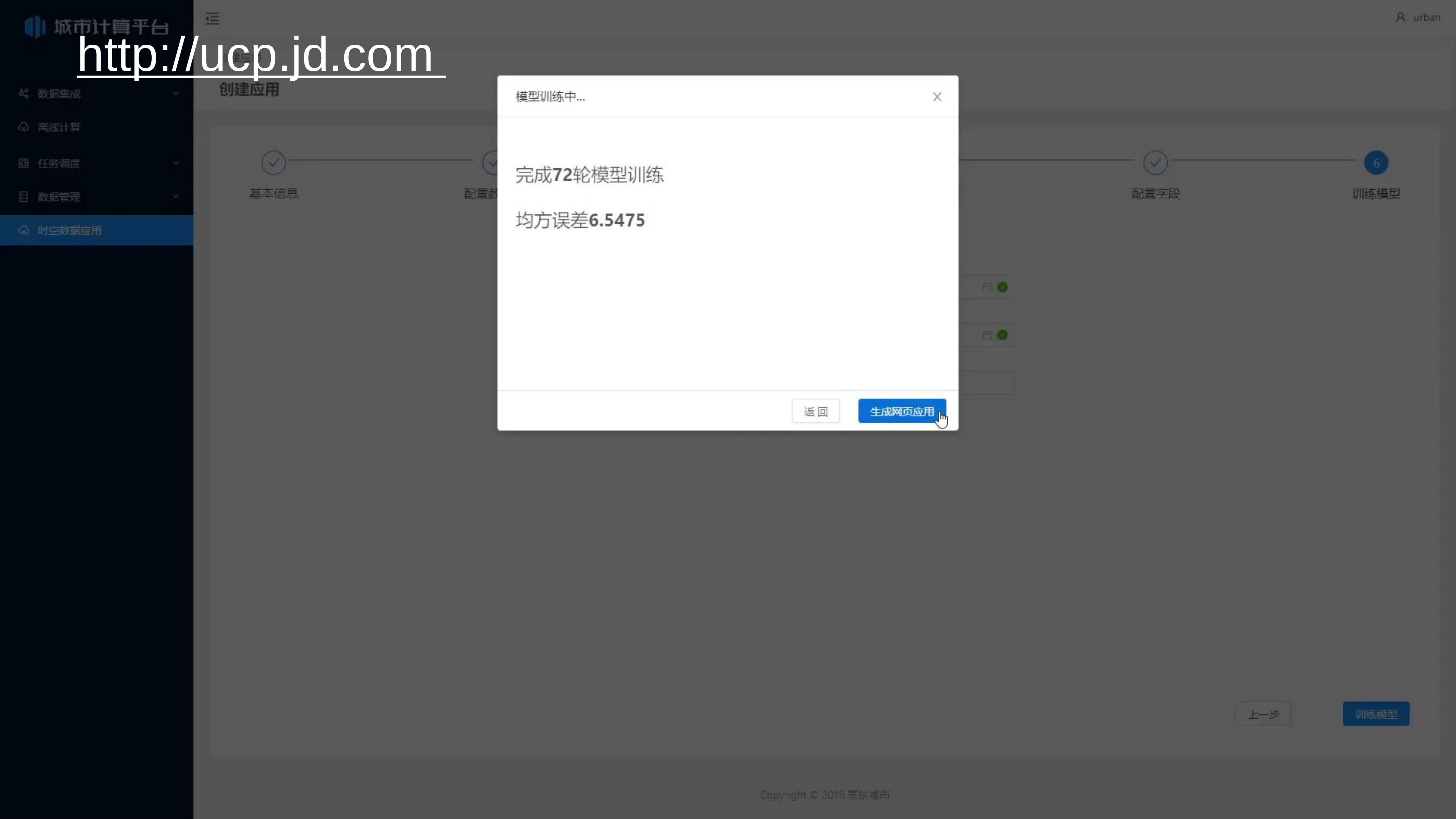
http://ucp.jd.com



The Platform adopts an open structure providing enterprise users



http://ucp.jd.com


http://ucp.jd.com
城市计算平台
urban
创建应用
数据集成
离线计算
任务调度
数据管理
时空数据应用
基本信息
配置字段
训练模型
6
模型训练中...
完成72轮模型训练
均方误差6.5475
返 回
生成网页应用
上一步
训练模型
Copyright © 2018 京东城市

Summary

- Framework of urban computing
- Research challenges, topics and techniques
- Unique spatio-temporal data
- Urban computing platform → City OS
- Many more challenges beyond technology…

A text book

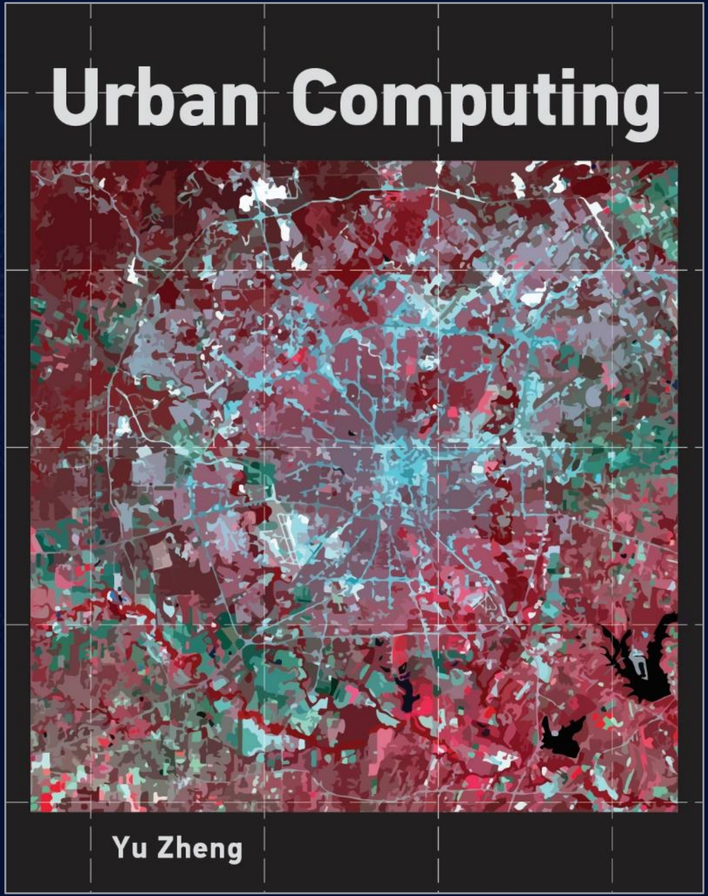
MIT Press

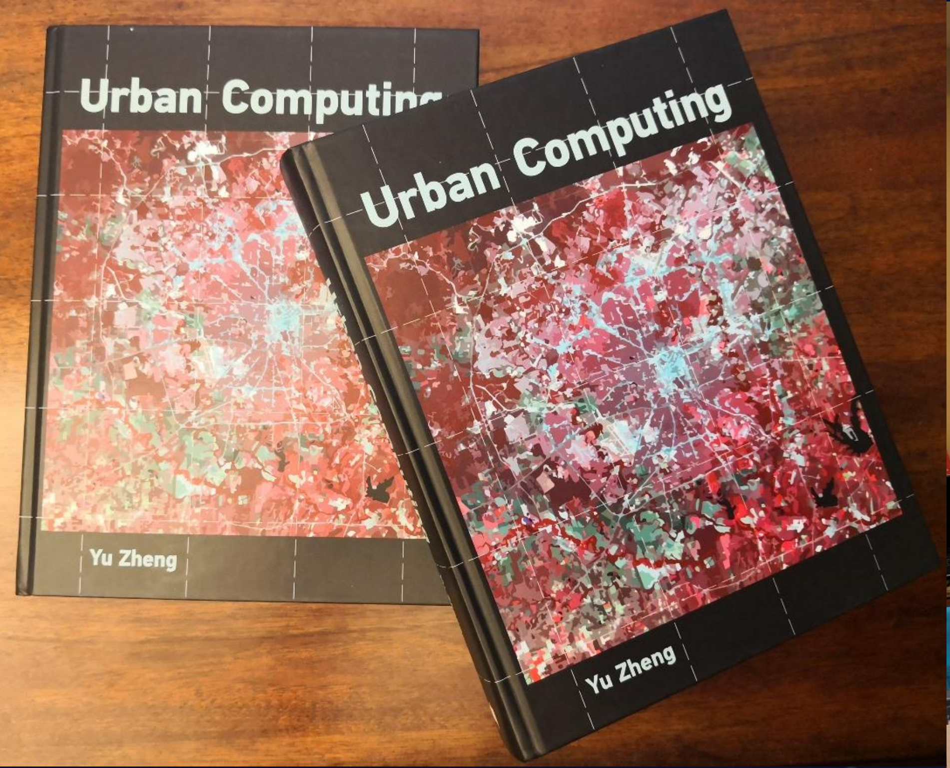
Thanks!

Dr. Yu Zheng, JD Digits

msyuzheng@outlook.com

http://urban-computing.com/yuzheng





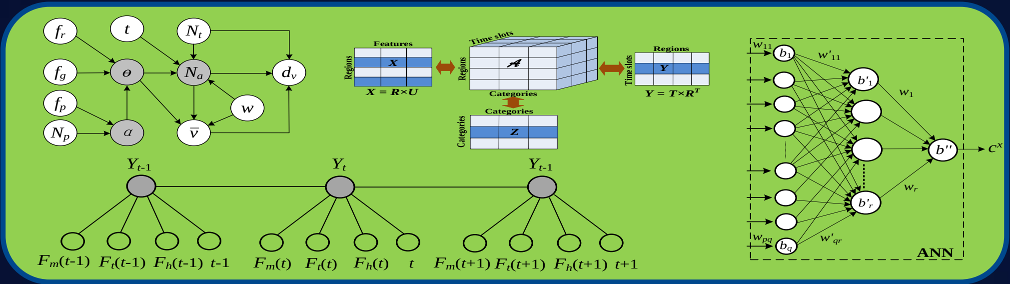
A text book

MIT Press

Data Scientist

京东城市
JD iCity

Models and Algorithms



Problems

Data