CS 671: Data-Driven Cyber-Physical Systems for Smart Cities

Desheng Zhang

Computer Science
Rutgers University

Spring 2017

New York City Taxi System
Who Am I

• Desheng Zhang
  • Assistant Professor
  • Department of Computer Science
  • Rutgers University
  • [https://www.cs.rutgers.edu/~dz220/](https://www.cs.rutgers.edu/~dz220/)
  • Office: CoRE 307
  • Phone: 848-445-8307
  • Email: d.z@rutgers.edu (including CS671 in your subject!)
My background

1. Bachelor & Master in CS at Heilongjiang University
2. Visiting Student at Shenzhen Institute of Advance Technology
3. Ph.D in Computer Science, University of Minnesota
4. Assistant Professor, Rutgers University
More about me

So, if I am not answering your email during a weekend, I am most likely...
About you

• Talk to your neighbors
  • Introduce yourself to him/her for two minutes
  • (maybe you could find your project partner)

• Introduce yourself to the class for 30 seconds
  • Undergrad Major
  • Degree Seeking
  • Research interests
Outline

• Logistics

• Course Structure

• Overview of Smart Cities Research

• Class Schedule
Class Website

• On Sakai
• https://www.cs.rutgers.edu/~dz220/CS671Spring17.html
About this class

• Lectures:
  • Thursdays: 3:20-6:20 pm

• Location:
  • Science and Engineering Resource Center (SEC) 206

• Office Hours
  • Mon 2:00-3:00pm by appointments at CoRE 307

• Teaching Assistant: Zhihan Fang
  • Email: zf72 AT cs.rutgers.edu
  • Office: CoRE 331
  • Office Hours: Wed 2-3pm
About this class

• Advanced Course on Smart Cities:
  • Reading
  • Presentations
  • Team Project

• Prerequisites:
  • Preliminary math knowledge
    • Calculus, Linear Algebra, & Probability
  • Skills for high-level programming languages are required.
    • C++, Java, R, Python or SAS

• No Textbooks are required:
  • links are provided
About this class

• Good for students who want to
  • Do Data-Driven **Research** on Data Science, Smart Cities, IoT
  • Have hands-on experiences about real data-driven projects
  • Improve their writing/presentation skills

• Not so good for students who want to
  • Practice Coding Skills
  • Do Hardware-related Projects
About this class

• Master Students:
  • This class **CAN** be used to satisfy **B requirements**
  • But each MSCS student should stop by the MSCS office (Hill 355) to make sure that his/her study plan is approved as per the graduation requirements.

• PhD Students:
  • This 67x class **CANNOT** be used to satisfy **B requirements**
  • Only Core Class CS 5xx can be used
  • Details on the CS website.
Outline

• Logistics

• Course Structure

• Overview of Smart Cities Research

• Class Schedule
Data-Driven Cyber-Physical Systems

Communication

Computation

Control

Cyber

Physical

Urban CPS
Urban Sustainability

The White House unveils new US$160 million Smart Cities Initiative

17th September 2015 Tom Teodorczuk
Urban Systems

NYC

Rome

Beijing

San Fran

D.C.

Shenzhen

Shanghai

Transportation

Telecom

Finance

Geography
Smart Cities Vision

Urban Systems

Data

Data-Driven CPS

Knowledge

Knowledge-Enabled Services

Feedback
Topics for 14 wks

• Introduction (1 wk)

• Basic Topics (7 wks)
  • Urban Sensing (1 wk)
  • Data Manage.&Processing (1 wk)
  • Data-Drive Modeling (3 wk)
  • Data Visualization (1 wk)
  • Data Predictive Control (1 wk)

• Proposal Presentation
  • 9th wk

• Special Topics (4 wks)
  • Novel Services (1 wk)
  • Dependency Analyses (1 wk)
  • Human-in-the-loop (1 wk)
  • Privacy and Security (1 wk)

• Final Presentation
  • 14th wk
<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topics and Reading Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jan 19</td>
<td>Intro (1 wk)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reading:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Urban Computing: Concepts, Methodologies, and Applications</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Cyber-Physical Systems: Executive Summary</td>
</tr>
<tr>
<td>2</td>
<td>Jan 26</td>
<td>Basic Topics (7 wks)</td>
</tr>
<tr>
<td>3</td>
<td>Feb 2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Feb 9</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Feb 16</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Feb 23</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Mar 2</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Mar 9</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Mar 23</td>
<td>Proposal (1 wk)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No Reading Assignment</td>
</tr>
<tr>
<td>10</td>
<td>Mar 30</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Apr 6</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Apr 13</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Apr 20</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Apr 27</td>
<td>Final (1 wk)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No Reading Assignment</td>
</tr>
<tr>
<td></td>
<td>May 4</td>
<td>Final project papers are due on May 4th 11:59PM EST.</td>
</tr>
</tbody>
</table>
Grades

• No Exams
• 15% for Class Participation
• 20% for Reading Summaries
• 20% for Topic Presentation
• 45% for Team-based Project
  • 10% for Proposal Report
  • 15% for Final Report
  • 20% for Proposal or Final Presentation

Individual Based 55%

Team Based 45%
Reading Summary (20%)

• Submitting 20 Reading Summaries

• Choosing 20 papers from assigned papers
  • Covering 10 out of 11 topics
  • 2 papers per topic
  • i.e., skipping one week

• A suggested format for summaries
Reading Summary: (20%)

• What is the key scientific question?
  • What it is **hard or important** to solve this question?

• What are existing approaches?
  • Why they are not sufficient?

• What the authors’ idea?
  • Why this is new?
  • Better than existing ones?
  • How they evaluate it?

• What are the strengths of this paper?
• What are the weaknesses of this paper?
• How you can address these flaws?
Topic Presentation (20%)

- Every student will be assigned with a topic
  - A paper (a set of papers) from Assigned Reading List
  - Your own work related to this topic

- Time 40 mins: 35 mins Talk & 5 mins Q&A

- Max 4 Presentations for Every Basic and Advanced Lecture

- Will send a poll to get five preferences of all students

- Coming to office hours before topic presentations
Team-based Project (45%)

• 10 Teams
  • 2-4 students per team
  • Assign based on topic interests (poll)
  • Form your own group if you really prefer

• Any new project related to smart cities
  • Common interests for all team members
  • Suitable scope for one semester
  • Be able to find data about it
  • Be able to evaluate it
Team-based Project (45%)

- Proposal Presentation (9\textsuperscript{th} wk)
  - 18 mins per team
  - Presented by first 2 team members

- Proposal Report (9\textsuperscript{th} wk)
  - 4 page-long double-column by ALL members

- Final Presentation (14\textsuperscript{th} wk)
  - 18 mins per team
  - Presented by the rest of team members

- Final Report (15\textsuperscript{th} wk)
  - 8 page-long double-column by ALL members

10\% 15\% 20\%
Before Any Presentation

• Any Presentation
  • Topic, Proposal, or Final

• Go to office hours:
  • Mon: 2-3pm: Desheng: CoRE 307
  • Wed: 2-3pm: Zhihan: CoRE 331

• Discussions
  • Background
  • Scopes
  • Technical contents
Bonus!

• Data-Driven Visualization

• Data-Driven Evaluation

• Demo: Animation or Video

• Present a set of Papers in Topic Presentation

• Present Your Own Work in Topic Presentation
Project Contest

• Vote by all students based on
  • Proposal Presentation
  • Final Presentation

• Three Teams will win prizes

• Announce after the final presentation
Suggestions about the class

• Suggestions are welcome
• Reading List
• Topics Interested
• Team Formation
Summary

• Come to all topic presentations (15%)
• Submit 20 summaries (20%)
• A Topic Presentation (20%)
• A team-based project proposal report (10%)
• A team-based project final report (15%)
• A team-based project presentation (20%)

Individual Based 55%
Team Based 45%
Useful References

• My website
  https://www.cs.rutgers.edu/~dz220/

• Mining of Massive Datasets - Stanford InfoLab

• Urban Computing at Microsoft Research
  https://www.microsoft.com/en-us/research/project/urban-computing/
Questions?
Outline

• Logistics

• Course Structure

• **Overview of Smart Cities Research**

• Class Schedule
Cross-Domain Cyber-Physical Systems for Smart Cities

Desheng Zhang

New York City Taxi System
Cyber-Physical Systems

Communication

Computation

Control

Cyber

Physical

Urban CPS
Why Urban?

Urbanization 2014

Traffic Congestion

The White House unveils new US$160 million Smart Cities Initiative

17th September 2015 Tom Teodorczuk

- $35M for Smart Cities Grants by NSF
  - $10M for CPS in 2016

- $70M for Transportation and Energy, by
  - DoT, DoE, NIST…
Urban Systems

NYC
San Fran
D.C.
Rome
Beijing
Shenzhen
Shanghai

Transportation
Telecom
Finance
Geography
Research Goals:

(1) **Understand** urban physical phenomena by *data-driven models*

(2) **Manage** urban physical systems by *knowledge-enabled services*
An Overview

Knowledge-Enabled Services

Data-Driven Models

Urban Systems

Cross-Domain Feedback

Ridesharing  
Improving Urban Phenomena  
*SenSys’13*

Human Mobility  
Understanding Urban Phenomena  
*MobiCom’14*

Cross-Domain Knowledge

Cross-Domain Data

Intellectual Core: Cross-Domain Systems
Goal

Modeling Human Mobility

Various Applications

Urban Scale & Real Time

© UMN
State of the Art

MobiCom’07
Zhang et al.

UbiComp’11
Lathia et al.

KDD’11
Cho et al.

UbiComp’13
Ganti et al.

Small Data Driven

Small Sample
Offline
Single Domain Driven

Biased Sampling & Overfitting
Opportunity: Cross-Domain Urban System Data

Modeling Human Mobility From a Cross-Domain Perspective
Contribution: Multi-view Bounding

Data-Driven Model-Integration Technique

- Considering **Domains** as Views
- Quantifying **Biases** of Individual Views
- Bounding **Range** of Unknown Phenomena by Interdependent Views

[MobiCom’14]
Bounding Range of Unknown Phenomena by Multi-Views with Quantified Biases

\[
\text{min}_{x^*, W} F(x^*, W) = \sum_{k=1}^{K} w_k \cdot D(x^*, x^k), \quad \text{s.t. } \delta(W) = 1
\]

High Penalty if a more-complete view deviates from the truth;
Low Penalty if a less-complete view deviates from the truth;

\[w^k: \text{Completeness Degree of a View } k\]
\[K: \text{Number of Views}\]
\[x^k: \text{Normalized Mobility Observed}\]
\[w^k: \text{Completeness Degree of a View } k\]
\[w = (w^1, \ldots, w^K)\]

Technique: Context-Aware Multi-View Bounding

Spatiotemporal Context: A → B during t

Mobility From A to B during t

\[x^*: \text{Mobility From A to B}\]

Bounding Range of Unknown Phenomena by Multi-Views with Quantified Biases

View Completeness Loss Function Regularization Function

Overall Weighted Deviation

High Penalty if a more-complete view deviates from the truth;
Low Penalty if a less-complete view deviates from the truth;
Vertex: Urban Region; Edge: Mobility Volume; Color: Urban Districts

Implementation: Region Level

Shenzhen 5-6 PM

Radiation 57%
Implementation: Street Level
Broader Impact: Open Data

20 GB Cross-Domain Data
Taxi; Bus; Truck; Subway; Cellphone; Smartcards;

Data Description for UrbanCPS
This 7GB dataset contains five kinds of data: Cellphone CDR Data, Smartcard Data, Taxicab GPS data, Bus GPS data and Truck GPS data of the Chinese City Shenzhen. This dataset is for academic research only. All rights reserved. For privacy concerns, all specific date info was removed and all identifiable IDs have been replaced by serial numbers in each kind of data.

1. CDR Data Format: 0055556100. 08:27:50,114.121305,22.57902
   SIM Card ID, Time, Latitude, Longitude;
   Download

   Smartcard ID, Time, Transaction type (21, 22, 31), Metro Station or Bus Line, Transaction Type;
   Download

   Download

...
Big Picture: Understanding Urban Phenomena for Smart Cities

Applications

- **Ride-sharing**
  - [SenSys’14]
- **Last-Mile Transit**
  - [IPSN’15]
- **Centralized Dispatching**
  - [ICCPS’15]
- **Distributed Navigation**
  - [RTSS’12]
- **Advertising**
  - [BigData’15]

Nonlinear Optimization with **Cross-Domain Knowledge**

Models

- **Human Mobility**
  - [Mobicom’14]
- **Traffic Speed**
  - [ICCPS’15]
- **Travel Pattern**
  - [SIGSPATIAL’15]
- **Passenger Demand**
  - [IPSN’15]
- **Transit Supply**
  - [BigData’13]

Multi-View Bounding with **Cross-Domain Data**

Systems

- Phone
- Credit card
- Car
- Train
- Bus
- Navigation
- Truck
- Bicycle
- Security camera

Cross-Domain Feedback
Interdependent Urban Phenomena

\[ \rho = \frac{\text{Mobility Demand}}{\text{Transit Supply}} \]

High Demand
Low Supply
Ridesharing

5PM-6PM
State of the Art & Limitations

1.3 Passenger Per Trip in NYC

70% Wasted

Similar Patterns

Sharing Capacity

T-Share (TKDE’14)

• Single-Domain Knowledge
• No Overall Urban Mobility Info

UberPool (2014.12)
Opportunity: Cross-Domain Knowledge

Traffic Speed
Transportation & Geography
Finance & Telecom

Mobility Demand
Telecom & Finance & Transportation

Minimizing Ridesharing Cost
Transportation: Mileage
Environment: Energy
Finance: Fare

cORide

Fare Model
Transportation & Finance

Transit Supply
Transportation & Geography
Contribution: coRide

- Cross-Domain Knowledge-Enabled Resource sharing Framework
- Bounded Online Approximation
- Real-world Implementation
- Multi-Scale Multi-Site Evaluation
- Potential Commercialization
Technique: 2 Approximation Algorithm

(1) Passenger Assignment with Minimum Spanning Tree (MST)

Mileage obtained by our *Online Approximation* is at most two times of the *Optimal Mileage*
Ridesharing Evaluation (1/4): Region

**Passenger Client**

**Onboard Device**

**coRide Server**

**Carpooling Vehicles**

**Subway Station**

**Station Exit Overpass**

**Central Control**
- CPU ARM CORTEX M3

**Motherboard**

**Prototype in Case**
- Onboard Sensing Camera
- Display
- ACC

**Reduced Mileage (%)**
(Gas and Congestion)

- 0%
- 20%
- 40%
- 60%
- 80%

- 1
- 5
- 9
- 13
- 17
- 21
- 25
- 29

49%
Ridesharing Evaluation (2/4): Urban

- Private Car: 24%
- Taxi: 37%
- Truck: 31%

Beijing
Ridesharing Evaluation (3/4): National

50K Commercial Vehicle Coverage

300K Private Vehicle Coverage

19%

29%
Ridesharing Evaluation (4/4): Worldwide

-- A Tale of Six Cities

New York City: 43%
San Francisco: 19%
Rome: 21%
Beijing: 37%
Shenzhen: 41%
Shanghai: 31%
Real World Impact 1: Dallas Taxi System

- Working with Mr. Vigil in Pchomes. Inc
- Practical Ridesharing in Dallas
Real World Impact 2: London Food Takeaway

A Ridesharing Problem for Food Takeout with Unique Challenges

• Traffic, Weather, Capacity, Timing, Food Types, Uncertain Origins
Big Picture: Improve Urban Efficiency with Sharing Economy

Contributions

• Resource Sharing Optimization Framework

• Cross-Domain Knowledge from Various Data-Driven Models

• Groundwork for Broader Logistics & Commercialization
Future Urban CPS Vision

3-5 Years
- Cross-Domain Modeling & Apps for Mobility
- Smart Cities Initiative
- Real-world Deployment for Research Impact

5-10 Years
- Data-Driven Physical Phenomenon Modeling & Apps
- Domains: Energy, Pollution, Safety, Privacy
- Homes, Buildings, and Cities, i.e., a Connected Smart World
Summary: Cross-Domain CPS for Smart Cities

Intellectual Core:
Cross-Domain Cyber-Physical Systems

Generalization:
Autonomous Driving;
Energy, Pollution, Safety;
Healthcare, Manufacturing

Cross-Domain Data
Multi-View Bounding
Real-world Implementation

Cross-Domain Knowledge
Optimization Framework
Potential Commercialization

Cross-Domain Data
Multi-View Bounding
Real-world Implementation

Models

Systems

Positive Feedback to
Improvement inSolutions

Critical Observation

Numerical Simulation
Bike Systems

New York City

Washington D.C.

San Francisco
Taxi Systems

NYC Taxicab Visualization

San Francisco Taxicab Visualization

Shanghai Taxicab Visualization

Rome Taxicab Visualization

Beijing Taxicab Visualization

Shenzhen Taxicab Visualization
Cellphone Mobility
Bus Mobility
Subway Mobility
Tiananmen Square

Oriental Pearl Tower

Window of the World

Beijing

Shanghai

Shenzhen

4th Ring Road

5th Ring Road

Outer Ring Expressway

G4 Expressway

National Heterogeneous Systems
Thanks

Data and More Work @ https://www.cs.rutgers.edu/~dz220/
Outline

• Logistics

• Course Structure

• Overview of Smart Cities Research

• Class Schedule
1. Overview

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topics and Reading Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jan 19</td>
<td>General Class Introduction</td>
</tr>
</tbody>
</table>

Suggested Reading:
- Urban Computing: Concepts, Methodologies, and Applications
- Cyber-Physical Systems: Executive Summary
- Research Directions for the Internet of Things
- Systems Computing Challenges in the Internet of Things
Urban Computing: Concepts, Methodologies, and Applications

YU ZHENG, Microsoft Research
LICIA CAPRA, University College London
OURI WOLFSON, University of Illinois at Chicago
HAI YANG, Hong Kong University of Science and Technology

(a) Motivation: Big cities, data and challenges

- Service Providing
  - Improve urban planning, Ease Traffic Congestion, Save Energy, Reduce Air Pollution, ...

- Urban Data Analytics
  - Data Mining, Machine Learning, Visualization

- Urban Data Management
  - Spatio-Temporal Index, Stream, Trajectory, and Graph Data Management

- Urban Sensing & Data Acquisition
  - Participatory Sensing, Crowd Sensing, Mobile Sensing

- Data Sources
  - Human mobility, Traffic, Air Quality, Meteorology, Social Media, Energy, Road Networks, POIs
2. Urban Sensing

Urban Sensing

Assigned Reading:
- How Long to Wait?: Predicting Bus Arrival Time with Mobile Phone based Participatory Sensing
- ParkNet: Drive-by Sensing of Road-Side Parking Statistics
- VTrack: Accurate, Energy-aware Road Traffic Delay Estimation Using Mobile Phones
- Discovering Regions of Different Functions in a City Using Human Mobility and POIs

Suggested Reading:
- People-Centric Urban Sensing
ParkNet: Drive-by Sensing of Road-Side Parking Statistics

Suhas Mathur, Tong Jin, Nikhil Kasturirangan, Janani Chandrashekharan, Wenzhi Xue, Marco Gruteser, Wade Trappe
WINLAB, Rutgers University, 671 Route 1 South, North Brunswick, NJ, USA
{suhas, tongjin, lihkin, janani, wenzhi, trappe, gruteser}@winlab.rutgers.edu
3. Data Management & Processing

Data Management & Processing

Assigned Reading:
- Managing Massive Trajectories on the Cloud
- TrajStore: An Adaptive Storage System for Very Large Trajectory Data
- Naiad: A Timely Dataflow System
- Pyro: A Spatial-Temporal Big-Data Storage System

Suggested Reading:
- SpatialHadoop: A MapReduce Framework for Spatial Data
Managing Massive Trajectories on the Cloud

Urban Applications

Traffic Modeling

Air Quality: Inference, Prediction

Spatial Classifier

Temporal Classifier

Interface

Trajectory Queries
Map-Matching

Trajectory Data Management

Indexes

Spatial Index

Spatio-temporal Index

Cloud Storages

Azure Table

Azure Blob

Azure Redis

Figure 1: Motivation Scenarios.
4. Data-Driven Modeling: Human Mobility

Data-Driven Modeling: Human Mobility

Assigned Reading:
- Human Mobility Modeling at Metropolitan Scales
- Inferring Human Mobility Patterns from Taxicab Location Trace
- Reconstructing Individual Mobility from Smart Card Transactions: A Space Alignment Approach
- Mobility Modeling and Prediction in Bike-Sharing Systems

Suggested Reading:
- Trajectory Data Mining: An Overview
- A survey on Human Mobility and its Applications
- Human Mobility Characterization from Cellular Network Data
- A Review of Urban Computing for Mobile Phone Traces
Human Mobility Modeling at Metropolitan Scales

Sibren Isaacman*, Richard Becker†, Ramón Cáceres†, Margaret Martonosi*, James Rowland†, Alexander Varshavsky†, Walter Willinger†

*Princeton University, Princeton, NJ, USA †AT&T Labs, Florham Park, NJ, USA
isaacman@princeton.edu, {rab,ramon}@research.att.com, mrm@princeton.edu, {jrr,varshavsky,walter}@research.att.com
5. Data-Driven Modeling: Urban Phenomena

Assigned Reading:
- Real-Time Trip Information Service for a Large Taxi Fleet
- City-Scale Traffic Estimation from a Roving Sensor Network
- A Cost-Effective Recommender System for Taxi Drivers
- Inferring Gas Consumption and Pollution Emissions of Vehicles throughout a City
City-Scale Traffic Estimation from a Roving Sensor Network

Javed Aslam  
College of Computer and Information Science  
Northeastern University  
ja@ccs.neu.edu

Sejoon Lim  
CSAIL  
Massachusetts Institute of Technology  
sjlim@csail.mit.edu

Xinghao Pan  
DSO National Laboratories  
pxinghao@dsogroup.com  

Daniela Rus  
CSAIL  
Massachusetts Institute of Technology  
rus@csail.mit.edu

(a) 3~5 am  
(b) 5~7 am  
(c) 8~10 am

(d) 1~3 pm  
(e) 5~7 pm  
(f) 9~11 pm
6. Data-Driven Modeling: Data Fusion

Assigned Reading:
- Diagnosing New York City’s Noises with Ubiquitous Data
- U-Air: When Urban Air Quality Inference Meets Big Data
- Exploiting Geographic Dependencies for Real Estate Appraisal
- Transfer Knowledge between Cities

Suggested Reading:
- Methodologies for Cross-Domain Data Fusion: An Overviewa
Diagnosing New York City’s Noises with Ubiquitous Data

Yu Zheng¹, Tong Liu¹,², Yilun Wang¹, Yanmin Zhu², Yanchi Liu³, Eric Chang¹
¹Microsoft Research, Beijing, China
²Shanghai Jiao Tong University, Shanghai, China
³Information Systems Department, New Jersey Institute of Technology, Newark, NJ, United States
{yuzheng, v-tongli, v-yilwan, echang}@microsoft.com; yzhu@cs.sjtu.edu.cn; yanchilyc@gmail.com
7. Data Visualization

Data Visualization

Assigned Reading:
- Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips
- TelCoVis: Visual Exploration of Co-occurrence in Urban Human Mobility Based on Telco Data
- Origin-Destination Flow Data Smoothing and Mapping
- Stacking-Based Visualization of Trajectory Attribute Data

Suggested Reading:
- Visual Analytics in Urban Computing: An Overview
Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips
8. Data Predictive Control

Reading:
- Exploiting Heterogeneous Human Mobility Patterns for Intelligent Bus Routing
- Rebalancing Bike Sharing Systems: A Multi-source Data Smart Optimization
- T-Finder: A Recommender System for Finding Passengers and Vacant Taxis
- Taxi Dispatch with Real-Time Sensing Data in Metropolitan Areas
Rebalancing Bike Sharing Systems: A Multi-source Data Smart Optimization

Junming Liu¹, Leilei Sun², Weiwei Chen³, Hui Xiong¹
¹Management Science and Information Systems, Rutgers University, USA, {jl1433, hxiong}@rutgers.edu
²Institute of Systems Engineering, Dalian University of Technology, China, leisun@mail.dlut.edu.cn
³Supply Chain Management, Rutgers University, USA, wchen@business.rutgers.edu

(a) Station Target Distribution  (b) Route with outliers (VN = 8)  (c) Route without outliers (VN = 12)
9. Project Proposal Presentation

• After Spring Break

• No New Topics
10. Novel Services

Assigned Reading:
- Catch Me If You Can: Detecting Pickpocket Suspects from Large-Scale Transit Records
- A Taxi Driving Fraud Detection System
- CrowdAtlas: Self-Updating Maps for Cloud and Personal Use
- Growing the Charging Station Network for Electric Vehicles with Trajectory Data Analytics

Suggested Reading:
- Modeling and Probabilistic Reasoning of Population Evacuation During Large-scale Disaster
Catch Me If You Can: Detecting Pickpocket Suspects from Large-Scale Transit Records

(a) Example Activities
(b) Trips
(c) Transit Records

(a) all passengers
(b) visitors
(c) shoppers
(d) thieves
11. Conflict and Dependency Analyses

Assigned Reading:
- Detection of Runtime Conflicts among Services in Smart Cities
- DepSys: Dependency Aware Integration of Cyber-Physical Systems for Smart Homes
- Scalable Social Sensing of Interdependent Phenomena
- Catastrophic cascade of failures in interdependent networks

Suggested Reading:
- The fragility of interdependency
- Identifying, understanding, and analyzing critical infrastructure interdependencies
Detection of Runtime Conflicts among Services in Smart Cities

M. M

Public

Private

Actuators

Public Actuators

Private Actuators

Actions

Intervene

Watchdog: Decision Making (Opt.)

Conflicts

Smart Services

Static Schedule
- Bus
- Waste Manage
- Street Light
- ...

Dynamic
- Transportation
- Safety
- Emergency
- Energy
- ...

Commercial
- Taxi Service
- Food Delivery
- Package Delivery
- ...

Private
- Vehicles Navg.
- Houses Keeper
- Phone Apps
- ...

Access

Sensors

Public Sensors

Commercial Sensors

Private Sensors

Email: {meiyi, pre

J. Stankovic*

and Information Technology

iversity,

ed

neberg@eit.lth.se
12. Human-in-the-loop

Assigned Reading:
- Taxi Driving Behavior Analysis in Latent Vehicle-to-Vehicle Networks: A Social Influence Perspective
- Using Humans as Sensors: An Estimation-theoretic Perspective
- Human Mobility, Social Ties, and Link Prediction
- Friendship and Mobility: User Movement In Location-Based Social Networks
Taxi Driving Behavior Analysis in Latent Vehicle-to-Vehicle Networks: A Social Influence Perspective

Tong Xu¹, Hengshu Zhu², Xiangyu Zhao¹, Qi Liu¹
Hao Zhong³, Enhong Chen¹, Hui Xiong³
13. Privacy and Security

Assigned Reading:
- DP-WHERE: Differentially Private Modeling of Human Mobility
- Elastic Pathing: Your Speed is Enough to Track You
- VPriv: Protecting Privacy in Location-Based Vehicular Services
- ZUbears against ZLyfts Apocalypse: An Analysis Framework for DoS Attacks on Mobility Systems

Suggested Reading:
- Differentially Private Transit Data Publication: A Case Study on the Montreal Transportation System
- Anonymization of Location Data Does Not Work: A Large-Scale Measurement Study
Elastic Pathing: Your Speed is Enough to Track You

Xianyi Gao, Bernhard Firner, Shridatt Sugrim, Victor Kaiser-Pendergrast, Yulong Yang, Janne Lindqvist
Rutgers University

Ground Truth  Predicted Path

1 Mile
2 km

Latitude
14. Final Presentation

• No New Topics
Questions?
Details about Grades

• **individual-based component 55% in total:**
  - **Class Participation 15%:** Please show up at all presentations and actively participate in the discussion after the presentation.
  - **Reading Summary 20%:** For 11 regular lectures (2nd-8th week and 10th-13th week), please read the introduction sections of all 4 papers in this lecture, and then select 2 papers from them to read all sections and write 2 summaries for each paper in the format given in the lecture slide. You can skip two summaries without hurting your grade, i.e., submit 20 summaries in total. Summaries are due in the beginning of class.
  - **Topic Presentation 20%:** Present a paper related to a topic in smart cities from the paper list. A topic presentation consists of 35 min talk plus 5 min Q&A.

• **Team-based component 45% in total:**
  - **Proposal Report 10%:** 4-page report describing the background, the problem, the existing solutions, and the proposed solution.
  - **Final Report 15%:** 8-page report: including 4 proposal report plus your implementation, evaluation, and conclusion.
  - **Proposal or Final Presentation 20%:** A student will present either in the proposal presentation or final presentation from 15 to 20 mins based on overall team size.
Details about Topic Selection

• Every student submits 5 of their preferred topics in the descending order among the following 10 topics.

• We will assign papers to you based on the topics you picked. Please send them to Zhihan at zf72@cs.rutgers.edu by Jan 25.

<table>
<thead>
<tr>
<th>Data Management and Processing,</th>
<th>Data Predictive Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data-Driven Modeling: Human Mobility</td>
<td>Novel Services</td>
</tr>
<tr>
<td>Data-Driven Modeling: Urban Phenomena</td>
<td>Dependency Analyses</td>
</tr>
<tr>
<td>Data-Driven Modeling: Data Fusion</td>
<td>Human-in-the-loop</td>
</tr>
<tr>
<td>Data Visualization</td>
<td>Privacy and Security</td>
</tr>
</tbody>
</table>
Team Formation:

• You are encouraged to find your own team members and let us know **by Jan 30**.

• If you cannot find anyone from the class, we will form teams **by Feb 2** based on your interests in topic selection sent to us previously.

• It is preferred that we have a mix-and-match for PhD students, master students, and undergrads in every team. In general, no more than one Ph.D is in the same team.