Urban Computing
–Using Big Data to Solve Urban Challenges

Dr. Yu Zheng
Lead Researcher, Microsoft Research
Chair Professor at Shanghai Jiao Tong University

Big Challenges in Big Cities

- Traffic congestion
- Industrial pollution
- Power transmission infrastructure
- Urban development
- Waste management
Big Data in Cities
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, streaming, trajectory, and graph data management, ...

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

Tackle the Big challenges in Big cities using Big data!

When Urban Air Meets Big Data

KDD 2013

http://urbanair.msra.cn/
Background

- Air quality monitor stations

- Map of 50km x 40km area with monitoring stations labeled S1 to S18

- Photos of urban and suburban areas with haze and pollution
We do not really know the air quality of a location without a monitoring station!
Inferring **Real-Time** and **Fine-Grained** air quality throughout a city using **Big Data**

- Meteorology
- Traffic
- Human Mobility
- POIs
- Road networks

- Historical air quality data
- Real-time air quality reports

Semi-Supervised Learning Model

- Philosophy of the model
  - States of air quality
    - Temporal dependency in a location
    - Geo-correlation between locations
  - Generation of air pollutants
    - Emission from a location
    - Propagation among locations
  - Two sets of features
    - Spatially-related
    - Temporally-related

Evaluation

- Overall performance


U-Air: when urban air quality inference meets big data.

Public website: http://urbanair.msra.cn/
- Transferred to CityNext and Bing Map China
- Working with Chinese Ministry of Environmental Protection
- Forecasting air quality in the near future
- To identify the root cause of the air pollution
Diagnosing Urban Noises using Big Data

UbiComp 2014
Background

• Many cities suffer from noise pollutions
  – Traffic, loud music, construction, AC...
  – Compromise working efficiency
  – Reduce sleep quality
  – Impair both physical and mental health
  – ...

• Urban noise is difficult to model
  – Change over time very quickly
  – Vary by location significantly
  – Depends on sound levels and people’s tolerance
  – The composition of noises is hard to analyze

311 in NYC

• 311 Data
  – A platform for citizen’s non-emergent complaints
  – Associated with a location, timestamp, and a category
  – Human as a sensor → crowd sensing
  – Implies people’s reaction and tolerance to noises
Goal

- Reveal the noise situation of each region in each hour
  - A noise indicator denoting the noisy level
  - Composition of noises in each location
Inferring Gas Consumption and Pollution Emission of Vehicles throughout a City

KDD 2014
Questions

How many liters of gas have been consumed by the vehicles, in the entire city, in the past one hour?

What is the volume of PM2.5 that has been generated accordingly?
Goals

- Estimate the gas consumption and vehicle emissions
  - on *arbitrary* road segment
  - at any time intervals
  - using GPS trajectories of a sample of vehicles

Real-Time CO Emission

Citywide

Top 4%
Computing with Heterogeneous Data

Fusion of data
- Use relevant data: intuitive and measure the relevance
- User different data sources correctly, rather than equally

Fusion of different models
- Understanding the merit of different models
- Integration of different algorithms cross disciplines
- Make the model semi-interpretable
Take Away Messages

- **3B**: Big city, Big challenges, Big data
- **3M**: Data Management, Mining and Machine learning
- **3W**: Win-Win-Win: people, city, and the environment

3·BMW

*ACM transactions on Intelligent Systems and Technology.*
Search for “Urban Computing”

Thanks!

Yu Zheng

yuzheng@microsoft.com