1. INTRODUCTION

Introduction

- Urban is a complex giant system;
- Problem we faced: traffic congestion, air pollution; sustainable development
- Big data, from earth observation system to our daily lift, providing a new way to better understanding our world. Data is the new oil of the digital world.

2. BIG DATA IN URBAN STUDIES

3. LESSONS FROM BIG DATA

Goals for Sustainable Development

Sustainable Cities and Communities

Make cities inclusive, safe, resilient, and sustainable
Traditional Way to Model Our World

Geographical Information System (GIS)
Global positioning System (GPS)

New Technologies to Understand Our World

Cloud Computing (Brain)
Artificial Intelligence (thinking mode)
Social Media (social way in cyberspace)
Mobile Internet (neural network transmission)
IoT / Sensor Network (Sensory organs)

Urban Informatics

Urban Sensing
Urban data integration
Urban information representation
Urban spatial development
Urban development strategy
Urban Planning

Urban Computing
IT on urban data management
Data analysis and mining
Urban construction and operation
Development policy
Urban Management

Pan-space Urban Sensing

Pan-space Urban Sensing is a cross-over study investigating the use of digital sensors (including remote sensing, internet of things, wireless sensor network, citizen-sensing, etc.) to obtain, compute, and analyze urban dynamic environments to achieve more sustainable and efficient cities.

Cubic Urban Sensing Framework

Physical Urban Space
Sensing urban physical structure with remote sensing techniques

Human Activity Space
Sensing human activities or behaviors with spatio-temporal big data

Cyber Urban Space
Sensing urban in cyberspace with social media and mobile network

Feature Space
Virtual Space
Human Being
Social Activity
Economy
Physical Space
Environment
Transportation
Landuse

Scale Space
Building-Parcel-Block
Above-under ground
In-out door
Second-Hour-Day-Month
History-Today-Future

Mapping Space
Field
Point
Polyline
Polygion
Network
Volume
Virtual Reality

Urban space

Urban Sensing Framework

Urban Sensing
Urban data integration
Urban information representation
Urban spatial development
Urban development strategy
Urban Planning

Urban Computing
IT on urban data management
Data analysis and mining
Urban construction and operation
Development policy
Urban Management

Pan-space Urban Sensing

Pan-space Urban Sensing is a cross-over study investigating the use of digital sensors (including remote sensing, internet of things, wireless sensor network, citizen-sensing, etc.) to obtain, compute, and analyze urban dynamic environments to achieve more sustainable and efficient cities.
Sensing urban physical structure with remote sensing techniques
To get the basic information about our city, like land use/land cover, building, roads, greens, etc.

Sensing urban social-economic with census, IoT, and electronic business data
Using social-economic related geospatial big data to help better understanding our social and economic system.

Sensing urban in cyberspace with social media and mobile network
Understanding citizen behavior or urban environment through online social media and study the relationship between physical and cyber urban space.
Physical Urban Space
Model Dynamic Urban Structure

Data covers main cities in China. Pan-space (physical, human, social-economic, and cyber space) data were integrated or fused, including but not limited to:
- Urban fundamental GIS data
- Census, Economic Census
- 365 * 24 hours MP positioning data
- Social media checkins

Urban Organs
Inferring urban detailed land use functions through pan-space urban information with Tensor-based artificial intelligence algorithms.
- 20 classes
- Overall accuracy: 91%

Urban Cells
Understanding detailed urban structure and behavior through the point of interesting (POI) data.
Each point with different colors represent a urban unit (shop, school, hospital, or factory, etc.)

Urban Pulse
Sensing and forecasting urban traffic congestion with GPS equipped floating vehicles data in urban arteries.
Algorithm: Bayesian deep learning with spatio-temporal correlations
Forecasting accuracy: 85%
Using mobile phone positioning data, we map urban human dynamic at very fine spatio-temporal scales.

**Dynamic Population Distribution**

\[ P_x = f(Z_{m,x,y,n}) \cdot \mu_1 \cdot LU + \mu_2 \cdot S + \mu_3 \cdot R + \mu_4 \cdot TN + \ldots \]

**Urban Rhythm**

Dynamic population distribution in 100 m resolution was estimated by using census and multiple geospatial big data with hybrid area-to-point geostatistical technique.

**Chinese Dynamics**

Location Based Service (LBS) data from Tencent Big Data center:

- 800 million users/day
- 50 billion location requests/day

**Chinese Footprint**

Mapping Chinese footprint worldwide with 400 millions Weibo (equate to Twitter) checkins to answer the question of "Where are the Chinese?"
Social media data representativeness?

The Chinese American estimation based on Weibo checkins was validated with US Census in county level.

Chinese American: (Based on 2010 US Census)

Chinese American: (Estimated based on Weibo checkins)

Model Chinese Mobility with Big Data

✓ Understanding individual human mobility is of fundamental importance for many applications from urban planning to disease spreading and traffic forecasting.

✓ We propose social media (Weibo) checkins as a proxy for human mobility, as it relies on publicly available data and provides high resolution positioning when users opt to geotag their posts with their current location.

Data source

Geo-computation

Pattern extraction

Mobility model

Application

Mobile phone data

Social media

Checkins

Floating car

Mobile

phone data

Internet

of-

thing

Traffic

OD matrix

ST pattern

spatial

interaction

Urban planning

disease

spreading

Human geography

Cross-validation

Floating population in 6th Census

Weibo estimated human mobility

Correlation: 0.85 (p-value < 0.001)

Understanding Chinese Mobility from Weibo

✓ Chinese mobility patterns is measured by big data computational strategy for identifying hundreds of millions of individuals’ space-time footprint trajectories.

✓ We discovered dialect-based culture ties control the Chinese mobility pattern.

✓ Our study provides solid evidence that Weibo checkins can indeed be a useful proxy for tracking and predicting human movement.

Understanding Chinese Mobility from Weibo

City-level Chinese mobility origin-destination (OD) matrix

City-level Chinese mobility in geographical space

Model Chinese Mobility

Cross-sectional OLS regression (in logs), city pairs

\[ \log(T_{od}/L_{od}) = \beta_0 \cdot \log(\text{Commuting}_{od}) \]

\[ + \beta_1 \cdot \log(\text{Dialect}_{od}) + F_1 \cdot F_2 \cdot \text{controls} + \epsilon_{od} \]

\(T_{od}/L_{od}\): mobility flows between city pairs/total Weibo user flows of the origin city

• [Commuting]: commuting distance\&time (pecuniary mobility costs)

• [Dialect]: dialect distance (non-pecuniary mobility costs)

• F: origin city fixed effect

• Fs: destination city fixed effect

• IV: historical dialect distance
Cyber Urban Space

Citizen Emotion Sensing and PM$_{2.5}$

China's high level of ambient air pollution causes sickness, excess mortality risk. Study health impact $\rightarrow$ Measure social cost?

Model global hourly PM$_{2.5}$ concentration with NASA MERRA2

Billions of geotagged Weibo posts

Control Variables: Weather, event, income, city property etc.

What a lovely weather! Let's go camping.

Today is heavily air polluted. My nose is stopped up.

Main findings:

- One standard deviation increase in the PM$_{2.5}$ concentration is associated with a 0.05-0.06 standard deviation decrease in the sentiment index.
- One standard deviation increase in the city's PM$_{2.5}$ concentration can be offset by a 6.5 thousand RMB ($940) increase in the city-level annual wage.

Air pollution impacts urban Chinese expressed sentiment, paper under review.

Emotion Sensing and Air Pollution

The Geography of Weibo posts, PM$_{2.5}$ concentration and sentiment index, and their national relationship.

Urban Data Scientist in Action

What can Urban Researcher Learn?

Lessons from Big Data
Urban Data Scientist

- Fundamentals
- Statistics
- Programming
- Machine Learning
- Text Mining
- Visualization
- Big Data
- Data Ingestion
- Data Munging
- Toolbox

Urban Data Scientist in Action

- Source Data
- Store Data
- Convert & ETL
- Transform Data
- Exploratory Analysis
- Model Build & Generate Insights
- Visualization
- Model Validation in Production

R in Action

Import

Transform

Visualise

Communicate

Python in Action

NumPy

SciPy

IPython Interactive Computing

pandas

matplotlib

Web data Scraping

Data Collection

Acquisition and Management
Remote, Social and Urban Sensing

<table>
<thead>
<tr>
<th>Organization</th>
<th>Top-down</th>
<th>Social Sensing</th>
<th>Urban Sensing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data accessibility</td>
<td>😞</td>
<td>😞</td>
<td>😞</td>
</tr>
<tr>
<td>Data representation</td>
<td>😞</td>
<td>😞</td>
<td>😞</td>
</tr>
<tr>
<td>Spatio-temporal granularity</td>
<td>😞</td>
<td>😞</td>
<td>😞</td>
</tr>
<tr>
<td>Model independence</td>
<td>😞</td>
<td>😞</td>
<td>😞</td>
</tr>
</tbody>
</table>

Small data and Big data

- **Wide data table with detailed attribute but limited samples**
  - Optimized samples are unbiased but with large error-variance
- **Long data table with limited attribute but massive samples**
  - Samples may biased but with small error-variance

Data fusion

- **Data fusion**

Data Analysis

Model, Computing and Analyst

Data Driven Modelling

- **Separate model**
  - Spatial analysis
  - Time series analysis
- **Spatio-temporal model**
  - Statistical model
  - Physical model for specific field
  - Statistical + physical model
- **Spatio-temporal data driven model**
  - Machine learning
  - Statistical deep learning
  - Artificial intelligence algorithms

Data Driven Modelling

- **Physical model**
- **Statistical model**
- **Full data driven model**

AI, ML, DL

- Caffe
- Chainer
- DL4J
- Keras
- MXNet
- Theano
- TensorFlow
Big Data in Earth Observation

Geospatial Analysis in the Cloud

Google Earth Engine
https://earthengine.google.com/

> 5 Petabytes of Earth observation data (imagery, weather, etc.)

Landsat
Sentinel
MODIS
ASTER
EO-1
DMSP-OLS

Geospatial Analysis in the Cloud

Global Forest Cover Change

Hansen et al. 2013. Science

Global Surface Water


This is the first map of forest change that is globally consistent and locally relevant. What would have taken a single computer 15 years to perform was completed in a matter of days using Google Earth Engine computing.

Professor Matt Hansen, University of Maryland

Analytic

Mapping and Analysis

D3.js
https://d3js.org/  https://www.jasondavies.com/
Cross-disciplinary cooperation!

Goodchild et al. (2012) PNAS:
“The supply of geographic information from satellite-based and ground-based sensors has expanded rapidly, encouraging belief in a new, fourth, or "big data," paradigm of science that emphasizes international collaboration, data-intensive analysis, huge computing resources, and high-end visualization.”

Thanks
Q & A

Collaborators: Yong Ge, Siqi Zheng, Wenjie Wu, Ying Long, Xingjian Liu, et al.
Acknowledgement: The National Natural Science Foundation of China

http://jianghao.wang
wangjh@lreis.ac.cn