清華大學

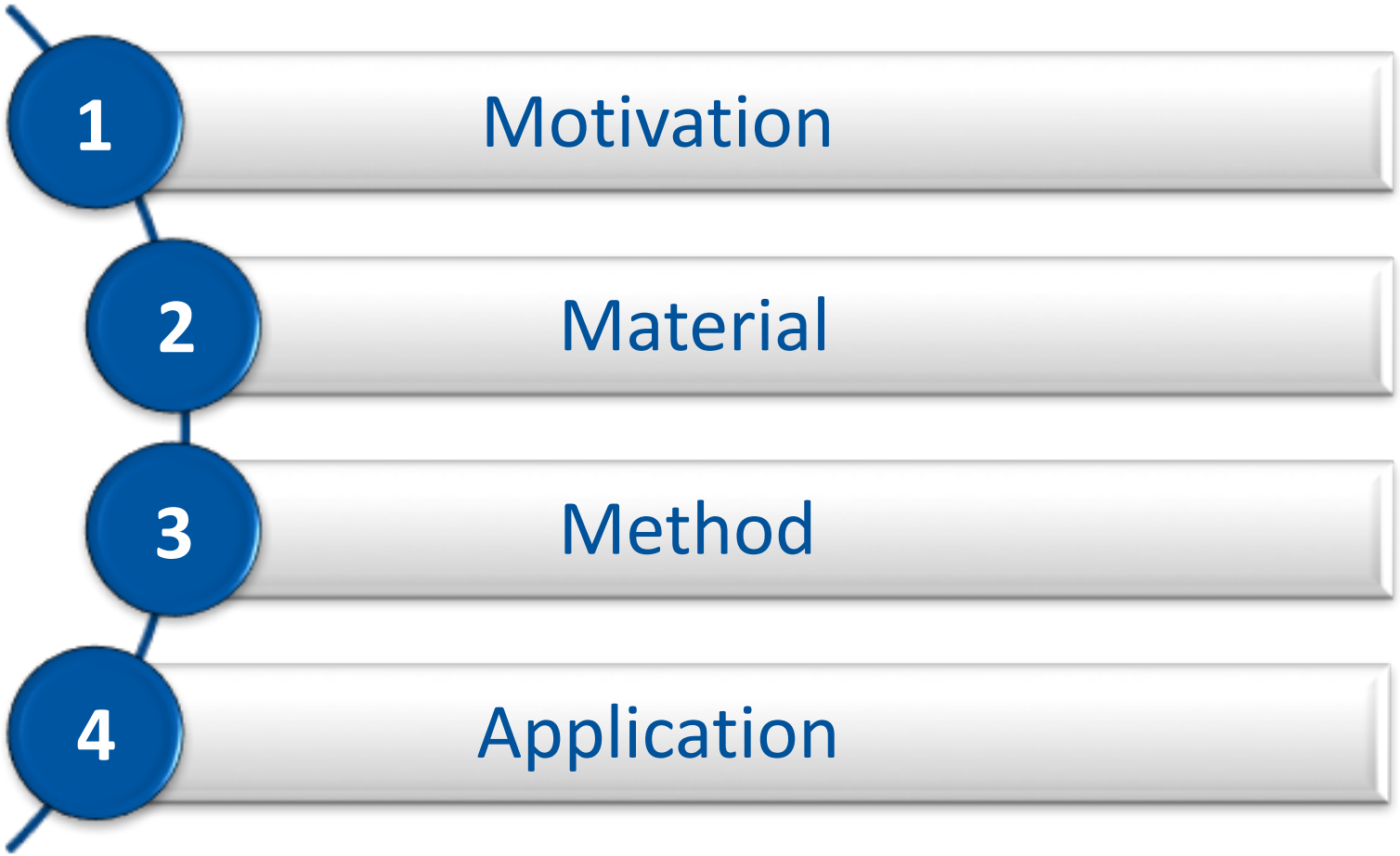
Tsinghua University

Context-aware Real-time Population Estimation for Metropolis

Fengli Xu

FIB, Tsinghua University

Outline



Motivation

Significance

The population distribution in cities during daytime.

Urban citizens contribute in cities daily lives.

Motivation

Problems of population census

- Very expensive
- High latency

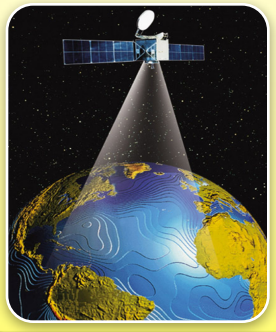
It’s not feasible to achieve real-time population distribution through census.

Motivation

Limitations of previous attempts

Call Detail Records[1]

- Low spatial resolution
- High latency(sparse records)



Remote Sensing Images[2]

- Require multiple datasets
- Can’t track day-time variation

1. Deville, et al. Dynamic population mapping using mobile phone data[J]. PNAS, 2014.
2. Stevens F R, et al. Disaggregating census data for population …[J]. PloS one, 2015.

Motivation

Limitations of previous attempts

Key points to address these problems

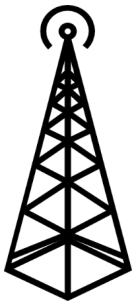
- **Appropriate datasets**
- **Advanced method**

Material

Cellular data access records

Device's ID||Start time||End time||BS ID||Location||Traffic volume

Large-scale



9600 BSs

150,000 users

Long duration

From August 1 to August 31, 2014

Fine-grained

Start and end time accurate to second

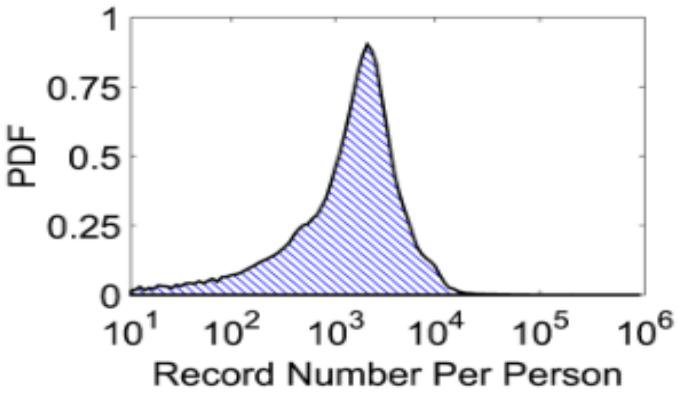
Contains 1.96 billion logs, total size over 300GB

We extract the number of access of each base station at granularity of one hour.

Material Dataset features

Extensive records

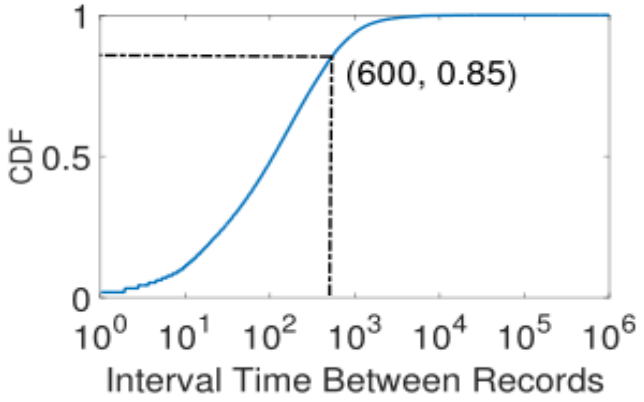
■ Most of users have more than 1,000 records in total.



High Sampling Rate

■ 85% of consecutive records happen in last than 10 Mins.

8.2 hours on average for call records [1]



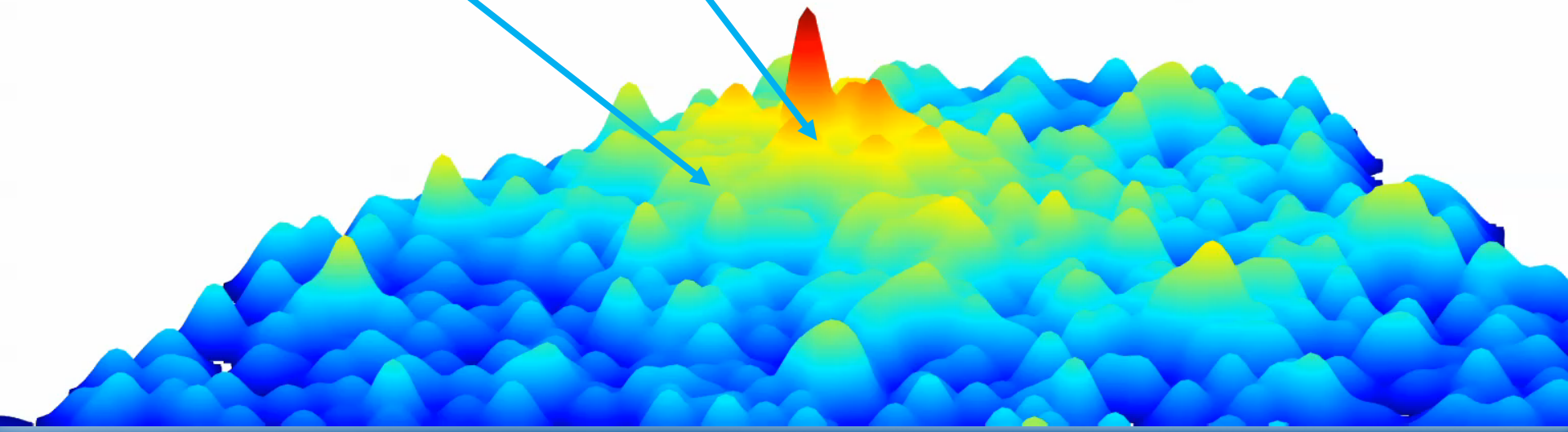
1. Gonzalez M C, et al. Understanding individual human mobility patterns[J]. Nature, 2008.

Material Visualization

Mobile users' behavior is related to the type of their location(physical context).

Can we achieve context-aware segmentation of the cities?

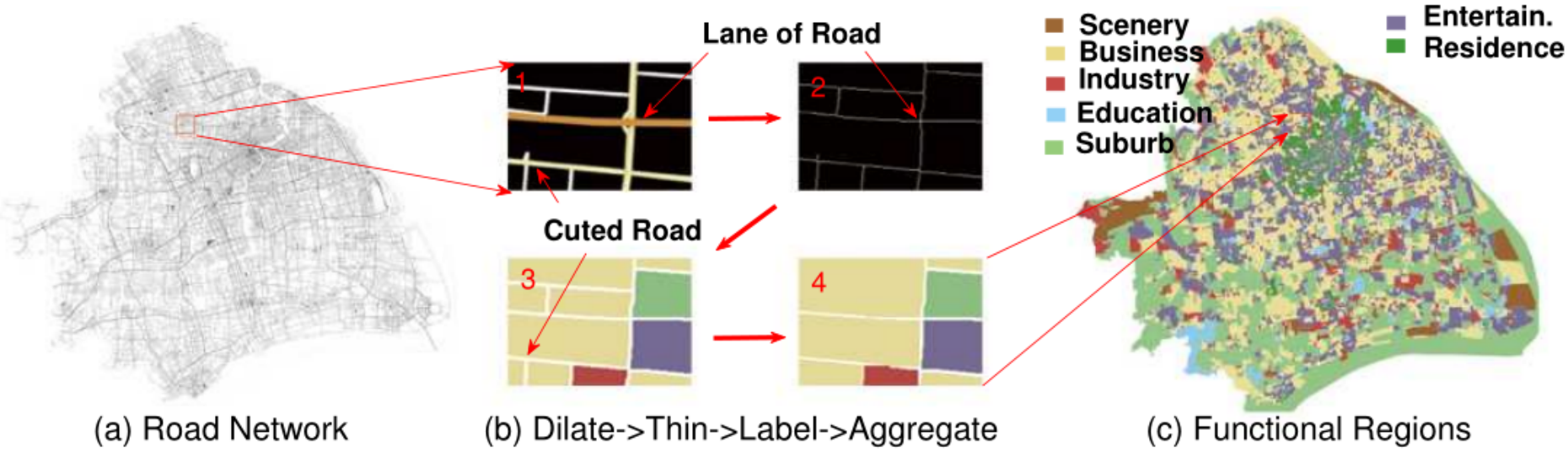
Different



Material

Context-aware segmentation

Road network forms a natural segmentation of urban environment. [1]



1. Yuan J, et al. Discovering regions of different functions …,SIGKDD, 2012.

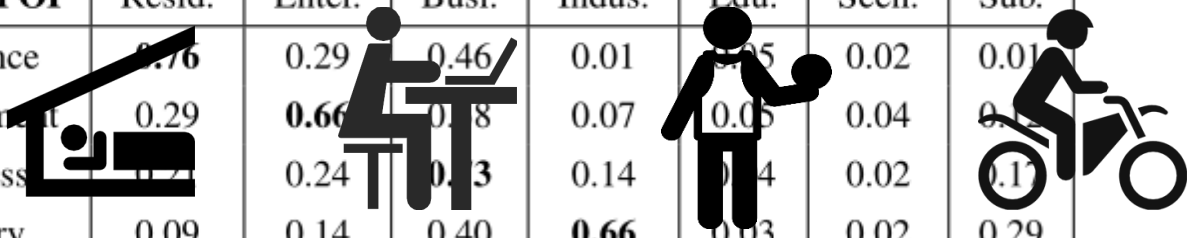
Material

Labeling the type of regions

POI — a specific point location of a certain function.

Region\POI	Resid.	Enter.	Busi.	Indus.	Edu.	Scen.	Sub.
Residence	**0.76**	0.29	0.46	0.01	0.05	0.02	0.01
Entertainment	0.29	**0.66**	0.28	0.07	0.05	0.04	0.11
Business	0.21	0.24	**0.43**	0.14	0.04	0.02	0.11
Industry	0.09	0.14	0.40	**0.66**	0.03	0.02	0.29
Education	0.14	0.22	0.22	0.08	**0.72**	0.03	0.17
Scenery spot	0.13	0.22	0.19	0.03	0.02	**0.77**	0.11
Suburb	0.06	0.08	0.17	0.10	0.02	0.02	**0.86**

Table 5. The mean value of TF-IDF vectors for every functional type.



1. http://map.baidu.com/

Material

Ground truth

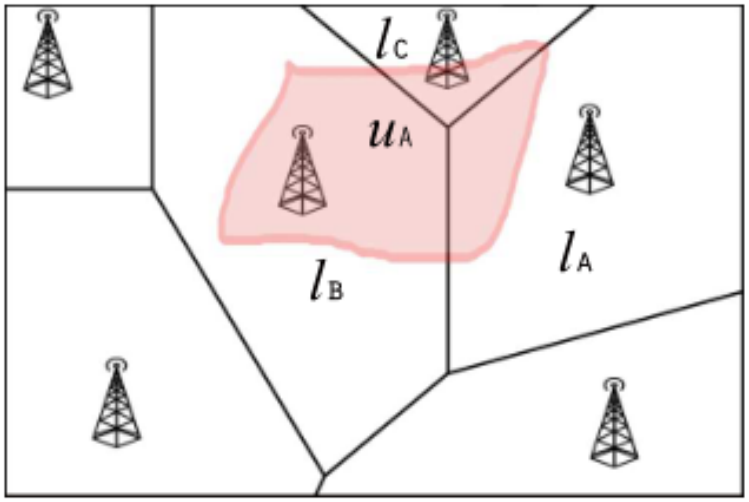
Worldpop project[1]:

- Only provide night-time population.
- Accurate(State of the art).
- High resolution 100mX100m.

1. http://www.worldpop.org.uk/.

Material

Data fusion



(a) Mapping of mobile users

(b) Mapping of Worldpop

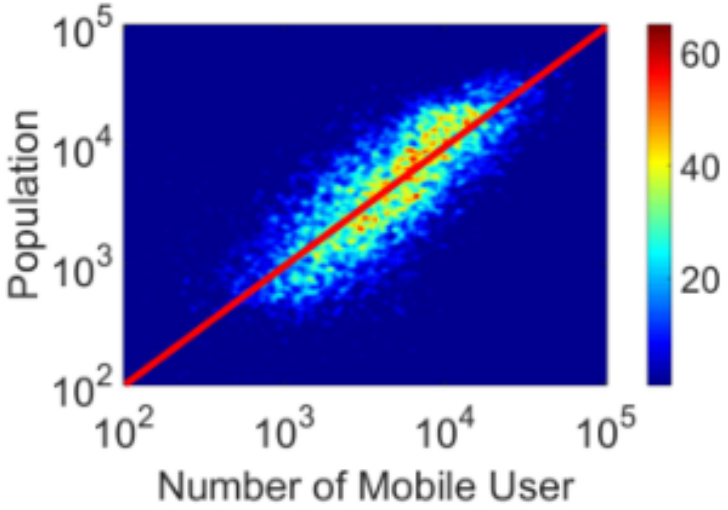
Mapping cellular data and worldpop data into segmented regions based on overlapping area.

Method Estimation model

- Inspiration: superlinear effect has been discovered in many fields in urban area, which is considered to be the result of intensive cooperation. [1]

$$\theta_u = \alpha(\rho_u)^{\beta}$$

$$\ln \theta_u = \ln \alpha + \beta \ln \rho_u$$



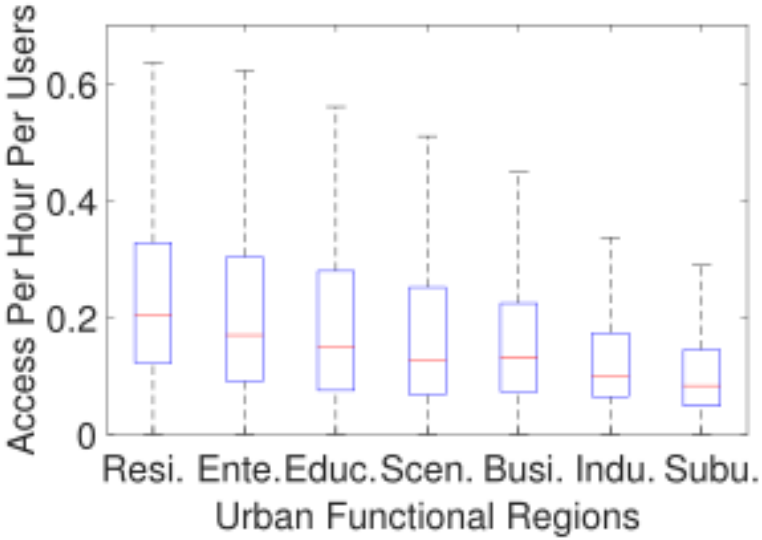
1. Bettencourt L M A. The origins of scaling in cities[J]. science, 2013.

Method Estimation model

- Context-aware estimation model.

$$\theta_u = \alpha_j (\rho_u)^{\beta_j}$$

$$j = 1, 2 \ldots 7$$



Users' behavior is spatial heterogeneous.

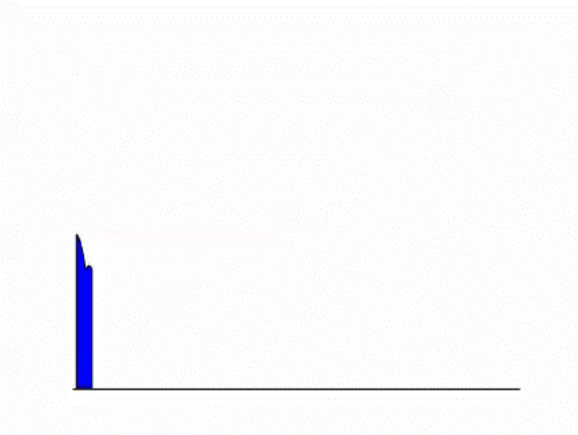
Method Estimation model

- Expand the model into a dynamic one.

$$R_t = \sum_u \theta_u \Big/ \sum \alpha_j (\rho_u)^{\beta_j}$$

$$\alpha_j^t = R_t \times \alpha_j$$

$$\hat{\theta}_u^t = \alpha_j^t (\rho_u^t)^{\beta_j}$$



α_j^t is scaled to model the temporal inhomogeneity of users' behavior, while β□j is fixed to model spatial characteristics.

Method

Evaluation method

Evaluation datasets

- Worldpop dataset.
- Transportation dataset(10 million taxi trips, 1 month)[1].

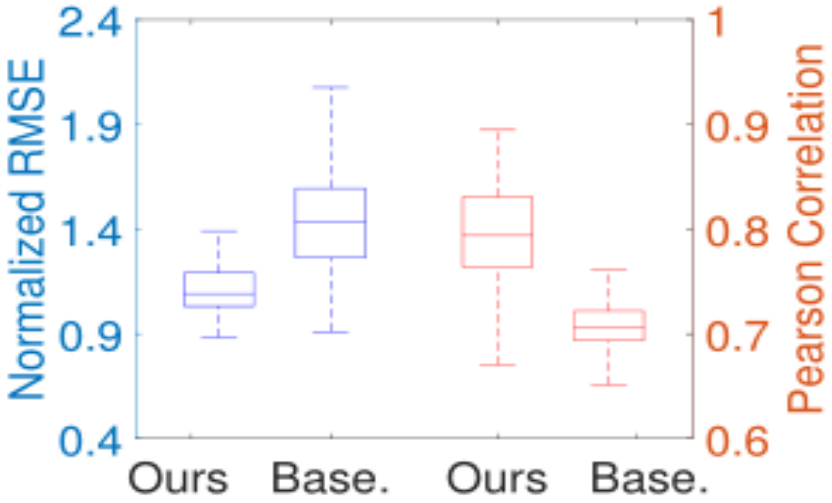
Evaluation schemes

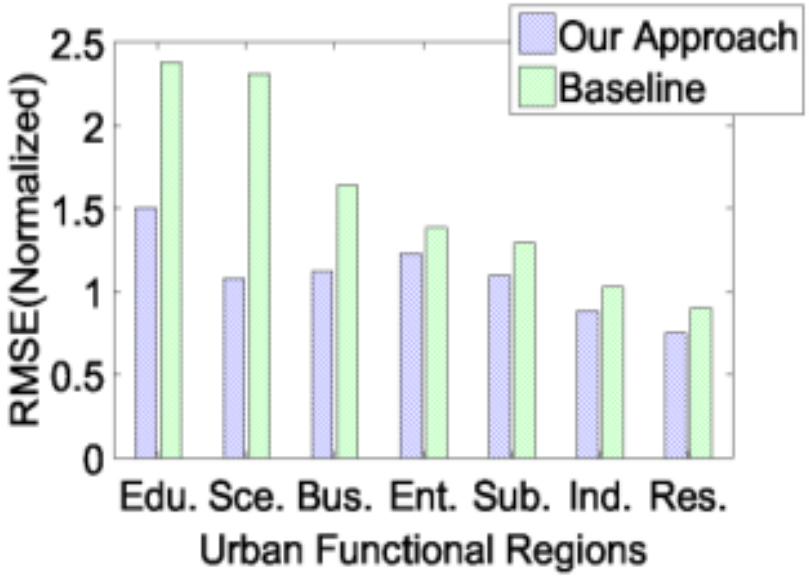
- Validate the night-time estimation with Worldpop data.
- Evaluate the real-time estimation with transportation dataset.

1. http://soda.datashanghai.gov.cn/.

Method

Evaluating night-time estimation

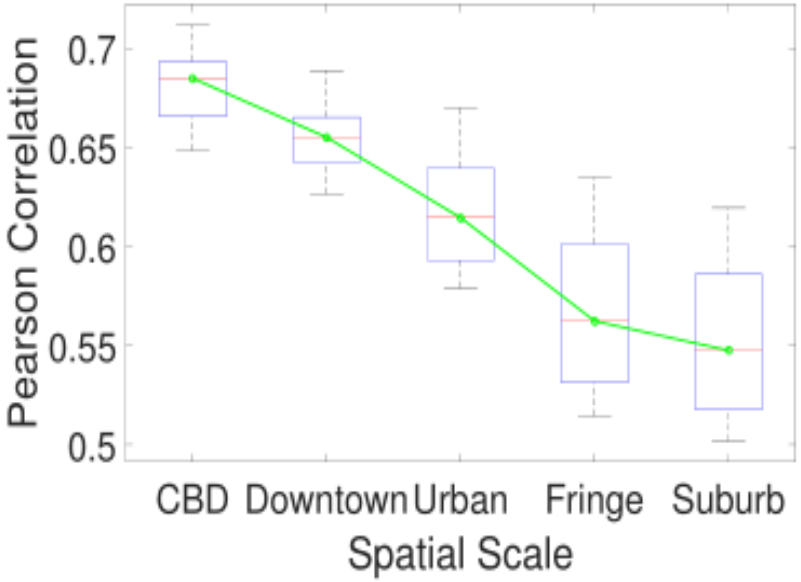


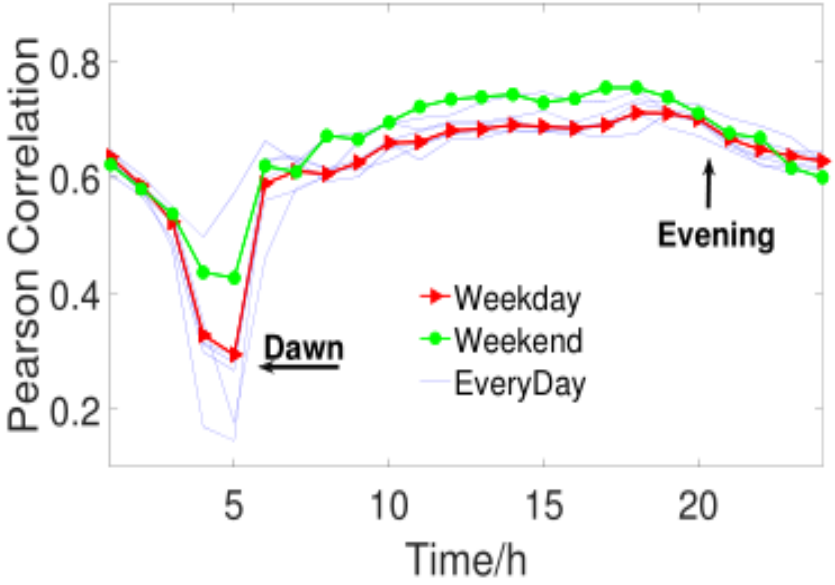


- Reduce 22.5% estimation error, enhance 12.5% correlation.
- Performance gain is most significant in education, scenery and business regions.

Method

Evaluating dynamic estimation



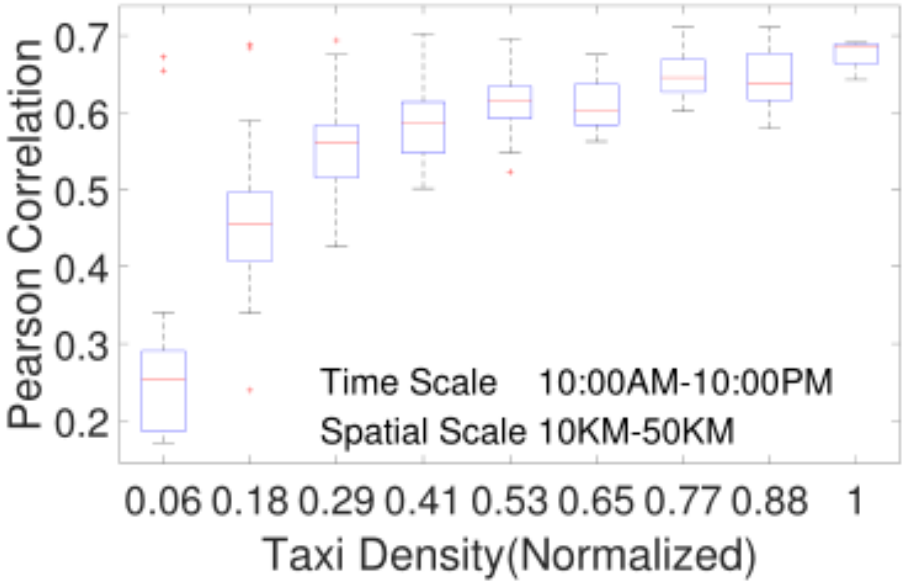


- Estimated population has a high correlation with taxi data in central area of urban.
- The correlation is significantly higher during day-time.

Method

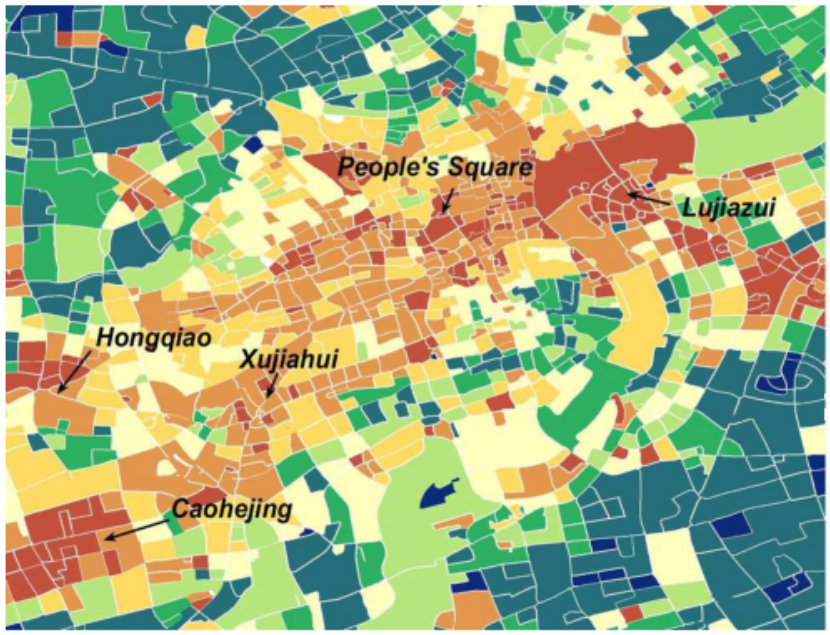
Evaluating dynamic estimation

The dynamic estimation matches well with taxi data.

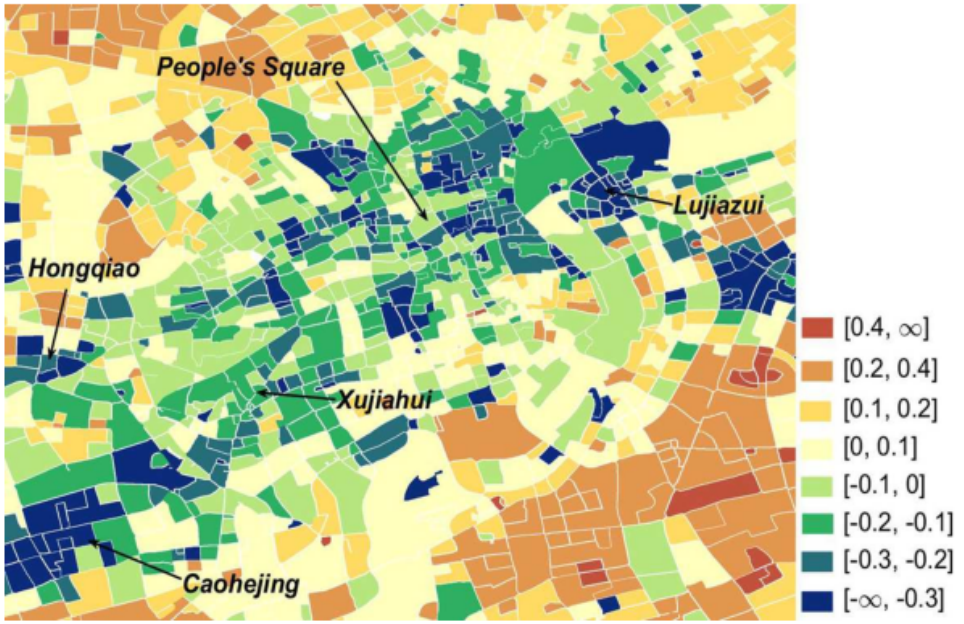


- The correlation monotonically increases with taxi density.
- Underlying reasons: taxi data can't capture population well when its density is low.

Application Observing urban dynamics



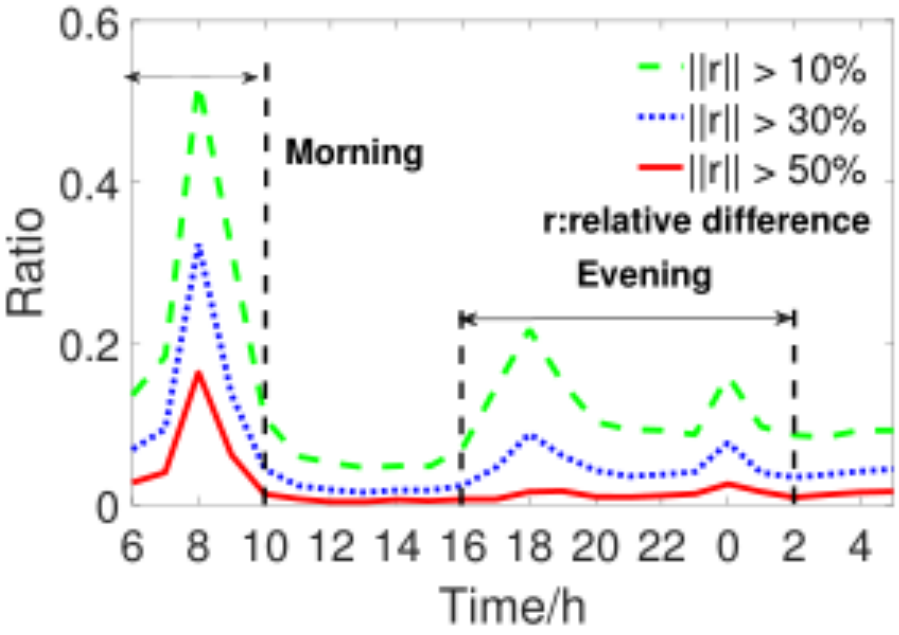
(a) Morning rush of downtown

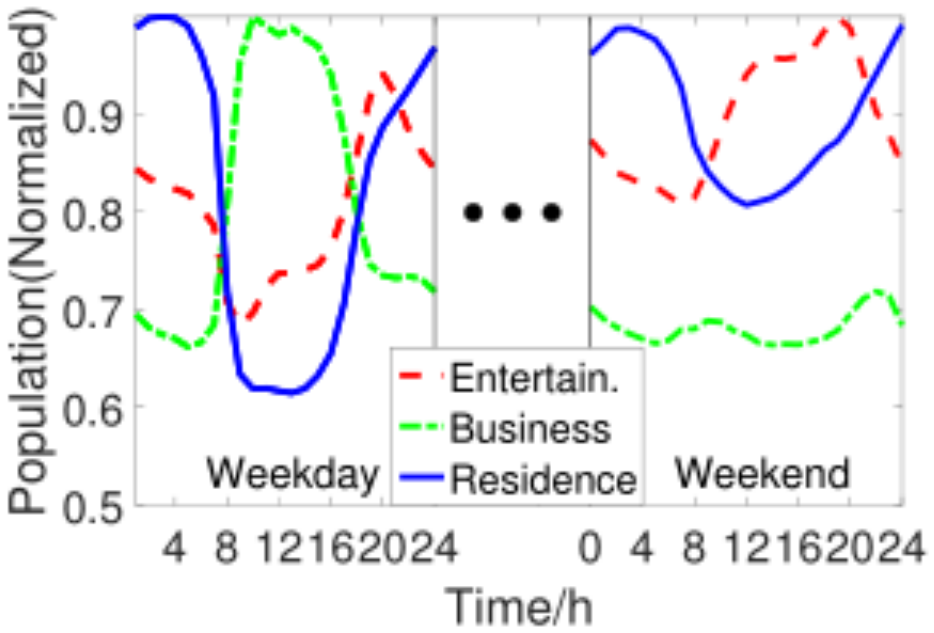


(b) Evening rush of downtown

Simple visualization can quantify the phenomenon of morning and evening rush.

Application Observing urban dynamics

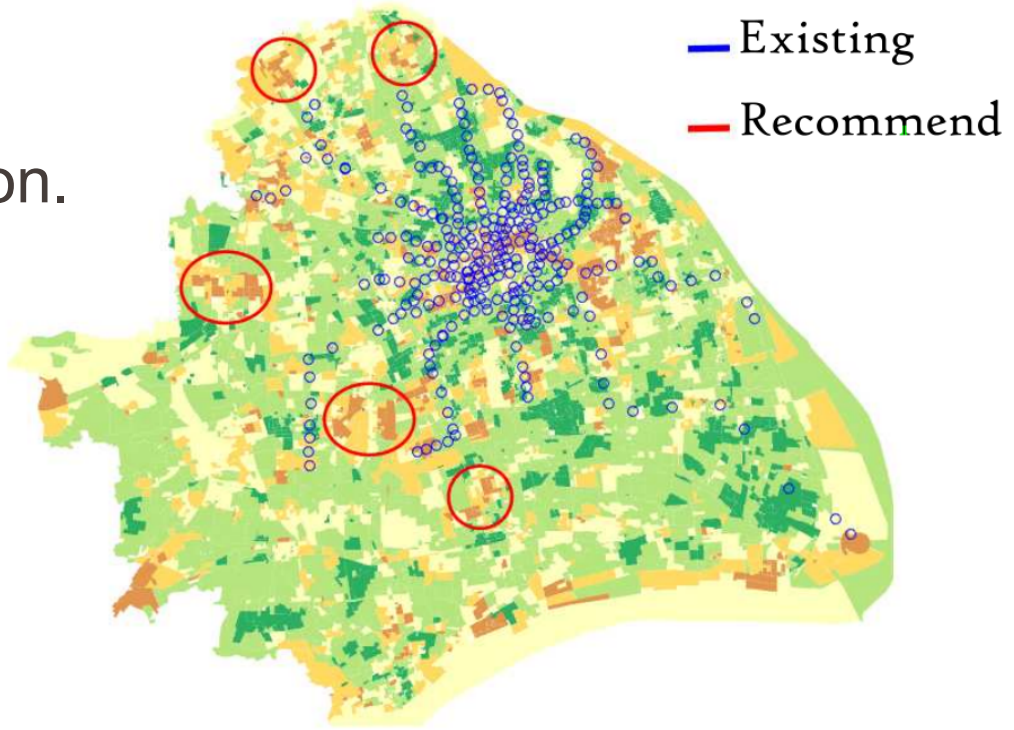




- Morning rush is more intensive than evening rush.
- Different functional regions have distinct population patterns.

Application Locate new subway station

- Warmer color represents higher variation of population.
- The regions with high population variation and no subway station are recommended identified.



Summary

- Appropriate data: collect 3G/LTE data access records of over 9,600 BSs with 150,000 subscribers for one month
- Advanced method: First estimation model to produce accurate real-time population estimation.
- Applications:
  - Visualizing and quantifying the dynamics of urban population.
  - Recommending locations for new subway stations.

Thanks you!

For Data Sample, Please Contact
xfl15@mails.tsinghua.edu.cn
liyong07@tsinghua.edu.cn
FIB-LAB: http://fi.ee.Tsinghua.edu.cn