CellTrans: Private Car or Public Transportation? Infer Users' Main Transportation Modes at Urban Scale with Cellular Data

Yi Zhao*, Xu Wang*, Jianbo Li†,

Desheng Zhang ‡ , Zheng Yang*

*Tsinghua University, †Qingdao University, ‡Rutgers University

清華大學

Tsinghua University

青島大学

QINGDAO UNIVERSITY

RUTGERS

Motivation

Understanding citizens' main transportation modes at urban scale is beneficial to a range of applications.

City Planning

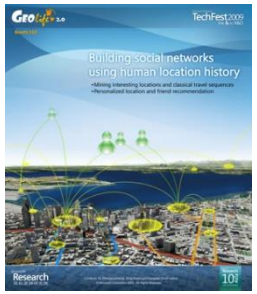
Transportation Management

LBS

Motivation

The inference of trajectory's transportation modes has been well-studied on GPS and phone sensor data, which are collected in a limited scale.

GPS Data



Geolife dataset[1]:
182 users

Sensor Data

UNIVERSITY OF SUSSEX

HUAWEI

SHL dataset[2]:
3 users

[1] Yu Zheng, Yukun Chen, Quannan Li, Xing Xie, and Wei-Ying Ma. 2010. Understanding Transportation Modes Based on GPS Data for Web Applications. ACM Trans. Web 4, 1, Article 1 (Jan. 2010),

[2] Lin Wang, Hristijan Gjoreskia, Kazuya Murao, Tsuyoshi Okita, and Daniel Roggen. 2018. Summary of the Sussex-Huawei Locomotion-Transportation Recognition Challenge. In Proceedings of UbiComp 2018.

Cellular networks

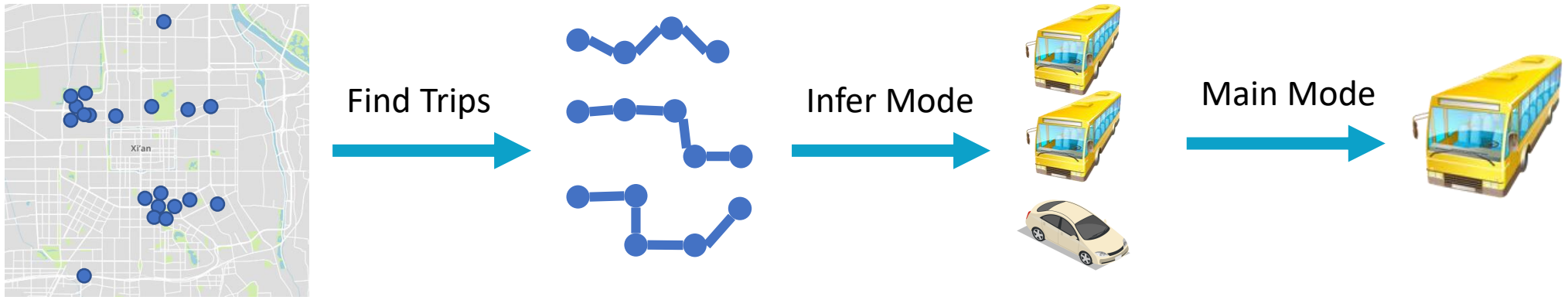
Fast development of cellular networks:

- Large scale, both spatially and temporally.
- Low cost, already collected for billing purposes.

Question

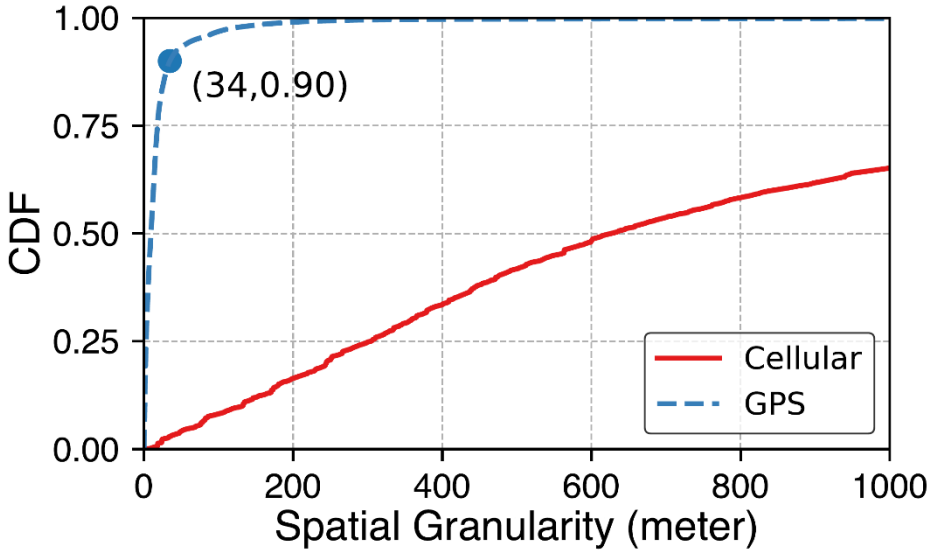
Can cellular data be used to infer users' main transportation modes?

- Direct solution based on previous methods:

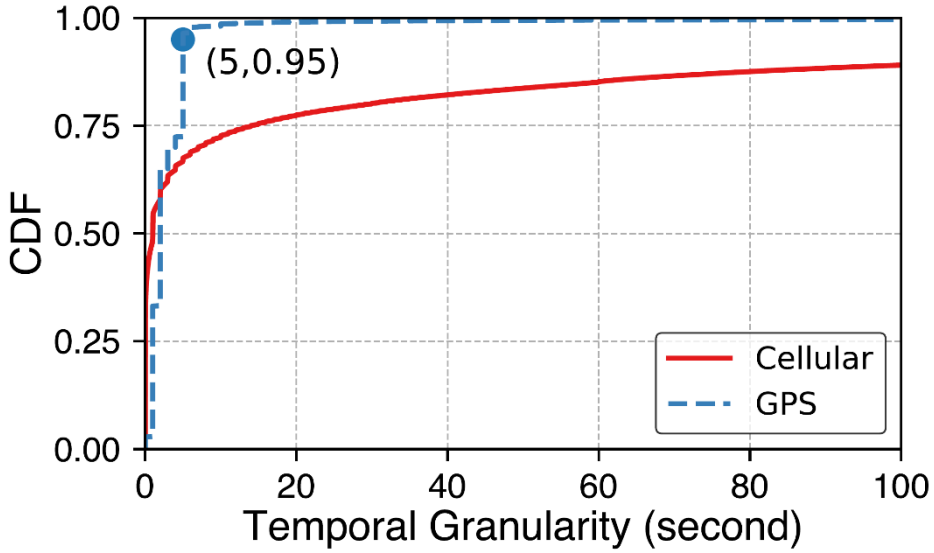


However, this direct solution does not work.

The direct solution does not work for cellular data:

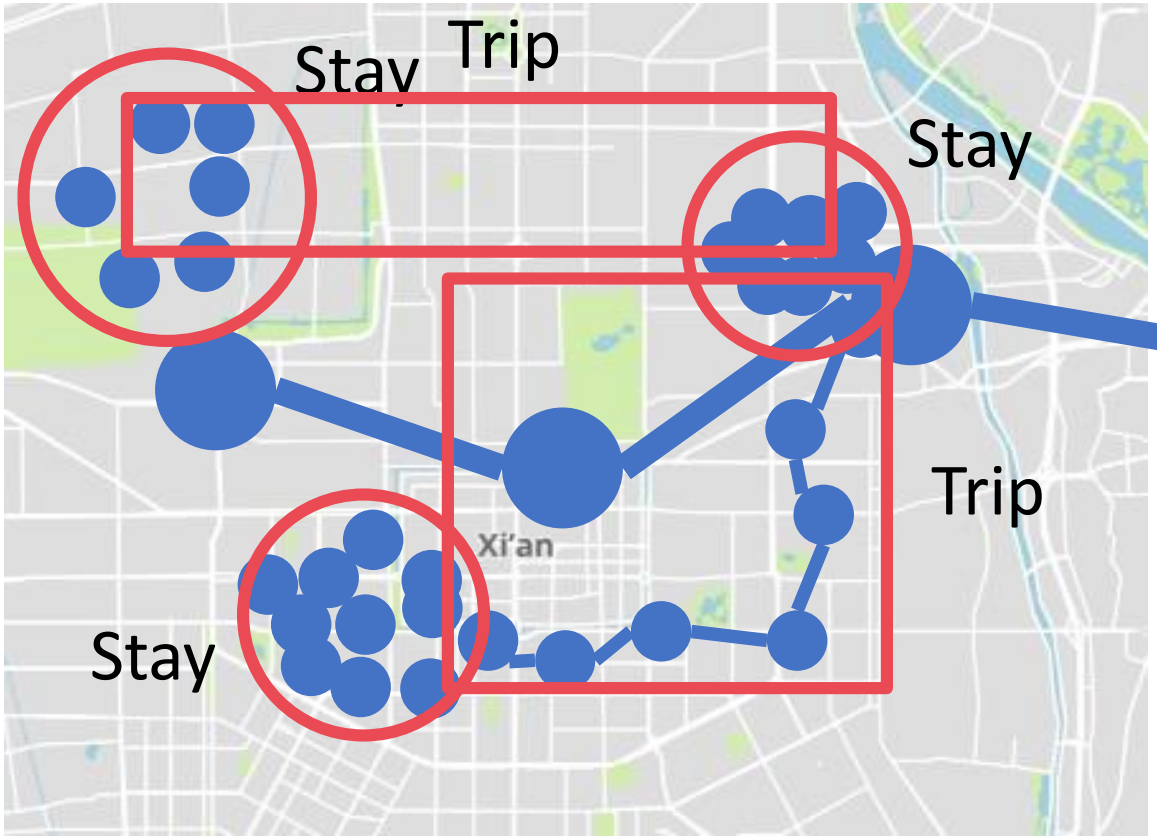


Coarse spatial granularity



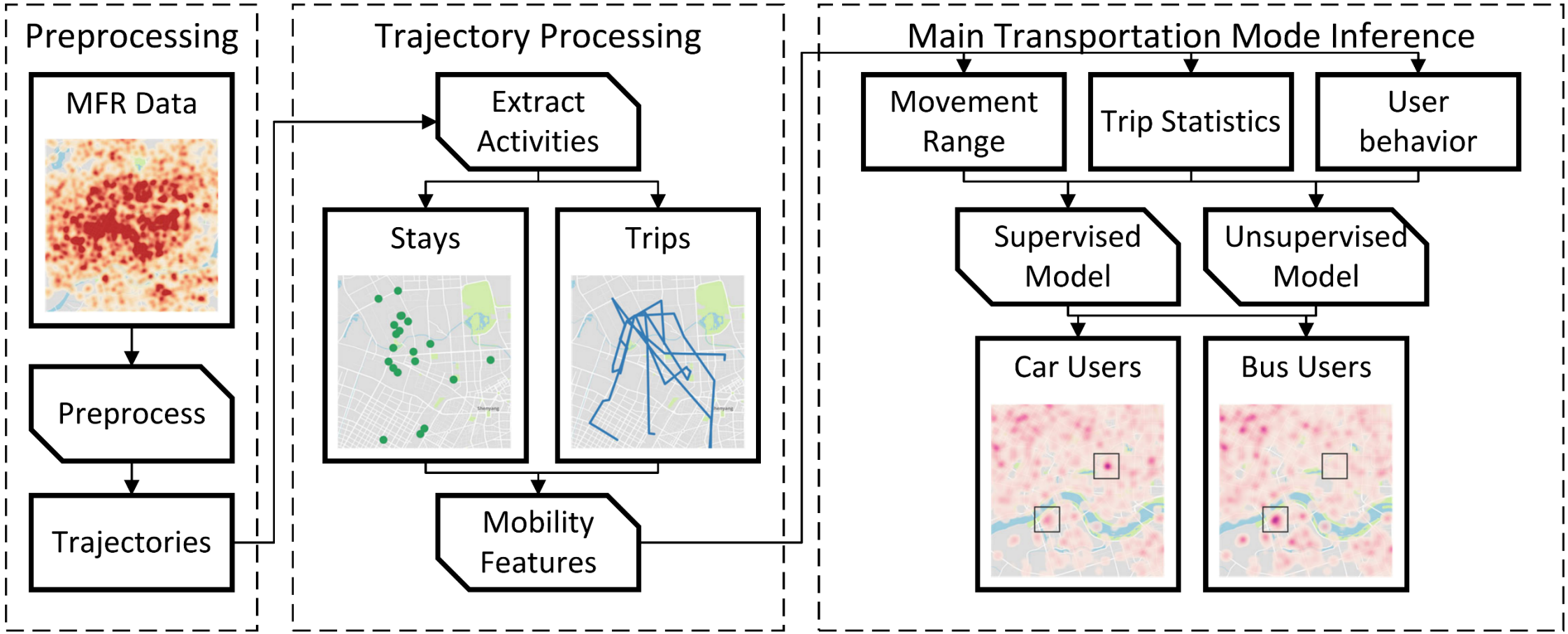
Irregular temporal sampling

CellTrans



- Instead of focusing on each trip, CellTrans considers a long period of users' location records.
- The expansion of observation time can compensate for the coarse spatiotemporal granularity of cellular data.

Framework of CellTrans



Dataset

We base our design on two large-scale cellular datasets from different cities: Shenyang and Dalian.

Statistics	Value
Records	8×10^9
Cell towers	1.2×10^4
Covered users	1.8×10^6
Covered area	$1.3 \times 10^4\,\mathrm{km}^2$
Covered period	Dec. 19, 2016 - Feb. 4, 2017

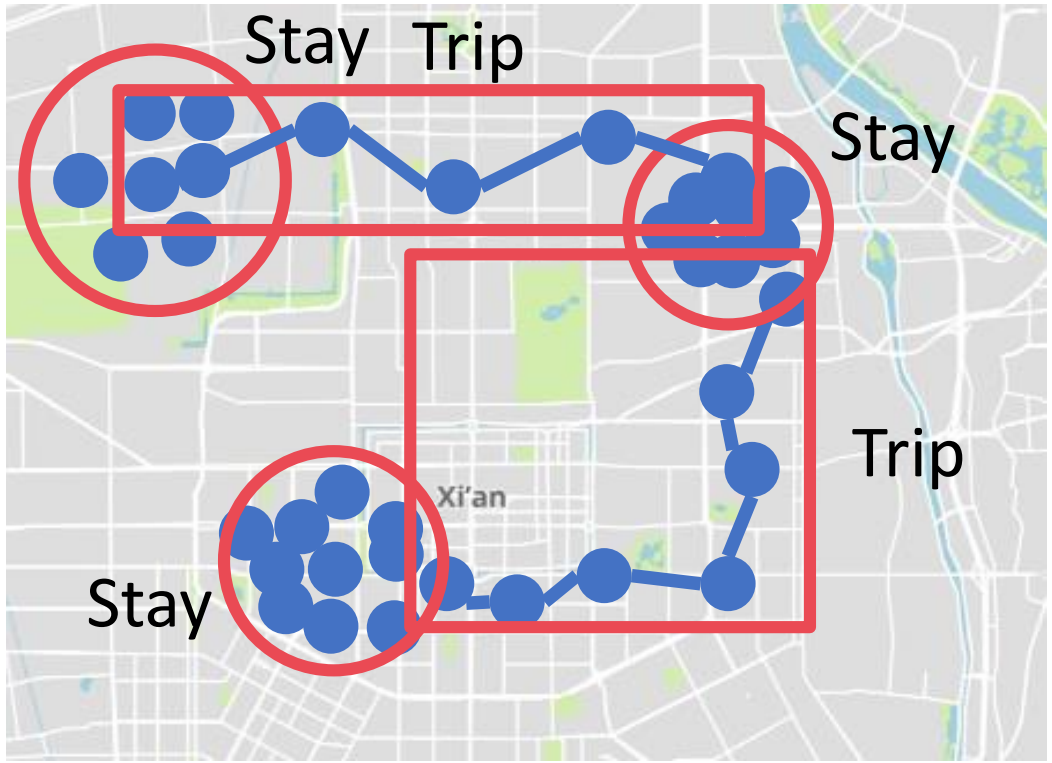
Shenyang

Statistics	Value
Records	12×10^9
Cell towers	1.2×10^4
Covered users	1.1×10^6
Covered area	$1.3 \times 10^4\,\mathrm{km}^2$
Covered period	Dec. 19, 2016 - Feb. 4, 2017

Dalian

Trajectory Processing

Parsing users' raw cellular data into stays and trips.[1]



- Stays usually correspond to users' activities like resting at home or working at office.
- Trips are trajectory segments when users travel from one stay region to another by some transportation means

[1] S. Jiang, J. Ferreira, and M. C. Gonzalez. 2017. Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore. IEEE Transactions on Big Data 3, 2 (June 2017), 208–219

Mobility Features

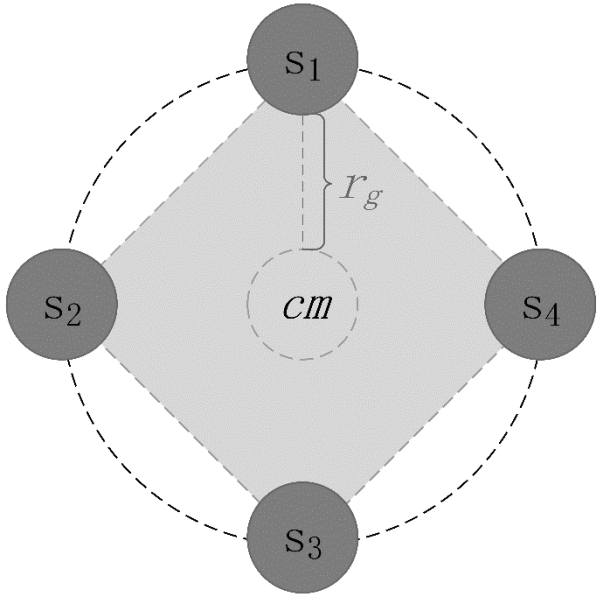
Movement Range

Trip Statistics

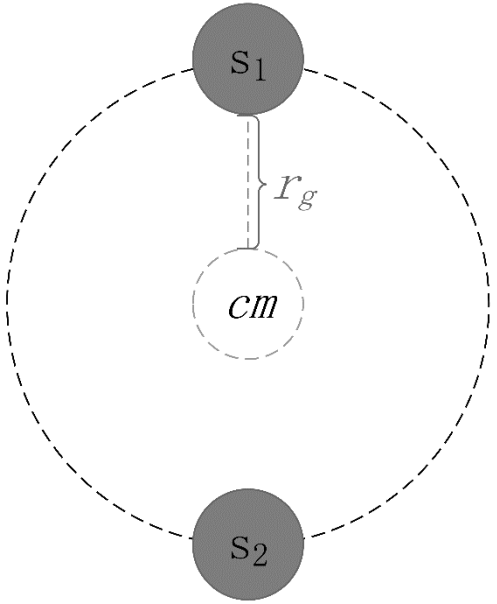
User Behavior

Mobility Features: Movement Range

It is easier for people driving car to visit more and further places compared to people taking public transportation.



$r_g=r_g$ $n_{cluster}=4$ $a=0.5*r_g*r_g$



$r_g=r_g$ $n_{cluster}=2$ $a=0$

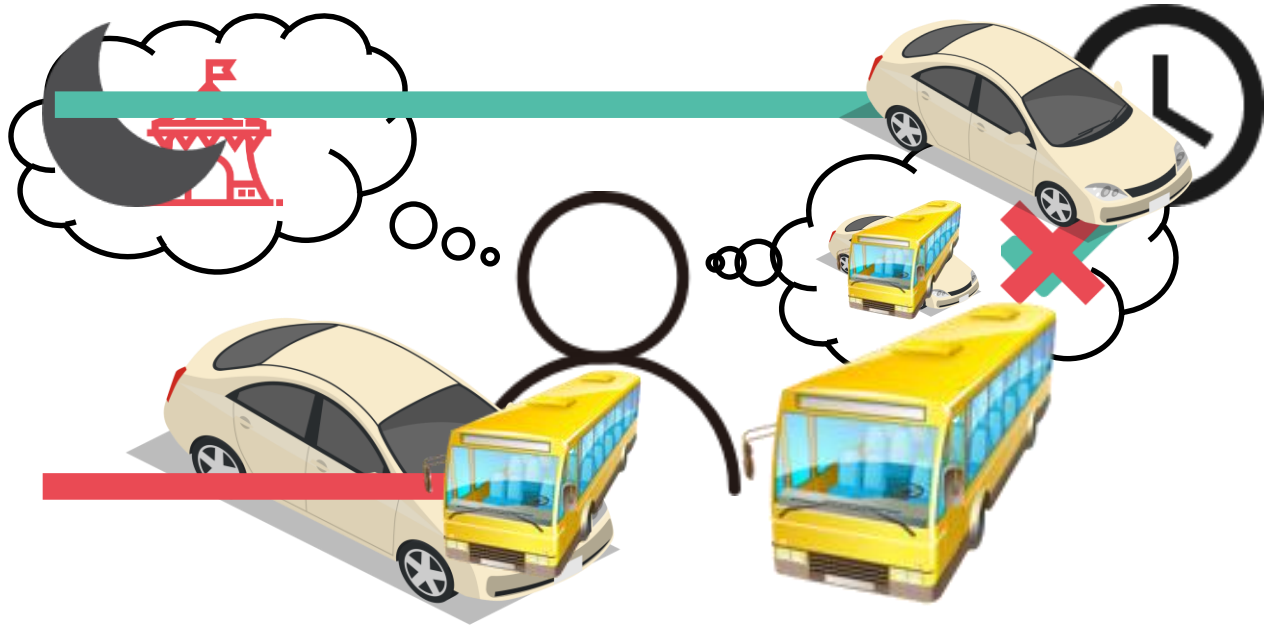
1. Radius of Gyration

2. # of Stay Clusters

3. Convex Hull Area

Mobility Features: Trip Statistics

The high-level statistics of trips can provide useful information to infer users' main transportation modes.



4. # of Trips

5. # of Night Trips

6. Average Speed

Mobility Features: User Behavior

The living pattern and economical status may be different between users of different modes.

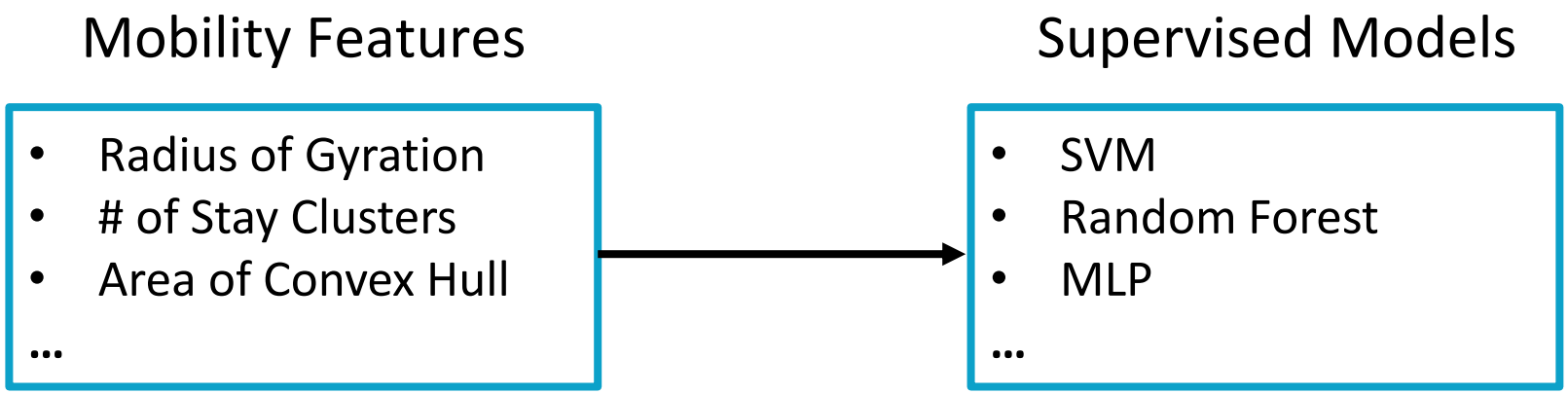
7. Network Access during Trip

8. Schedule

9. House Price

Mode Inference Model

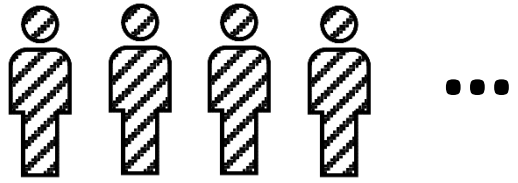
Scenario 1: With Labeled Users. We assume that partial users' actual modes are known, so a supervised model can be trained.



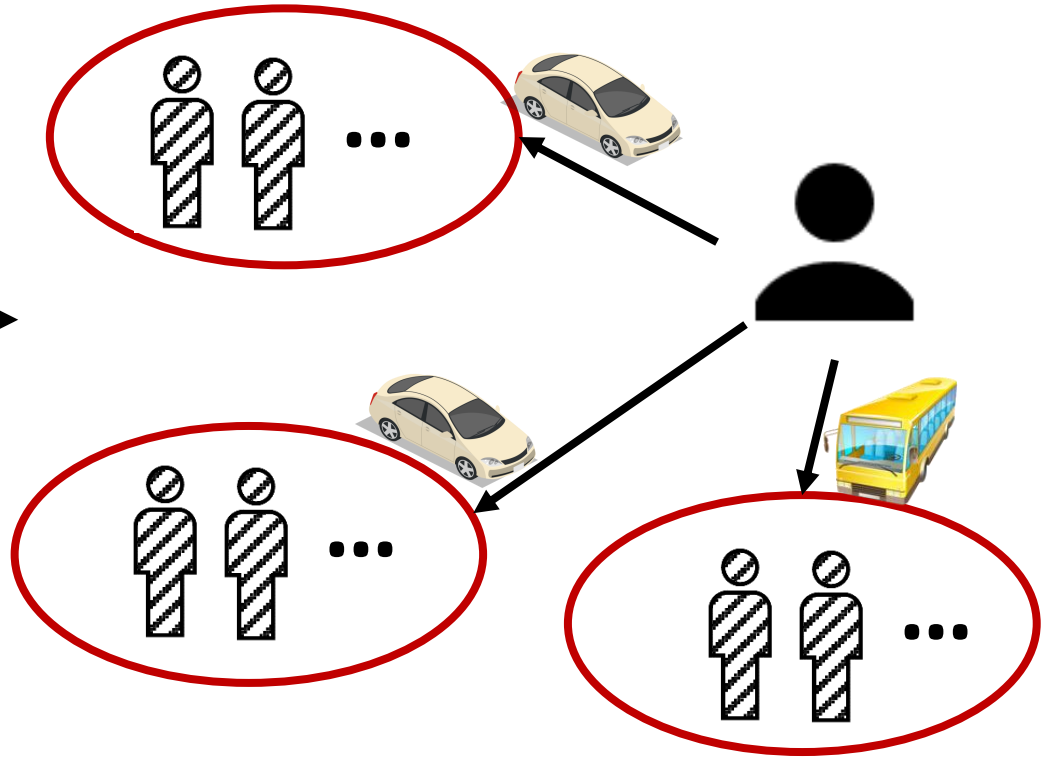
Mode Inference Model

Scenario 2: Without Labeled Users:

Car or Public trans. users

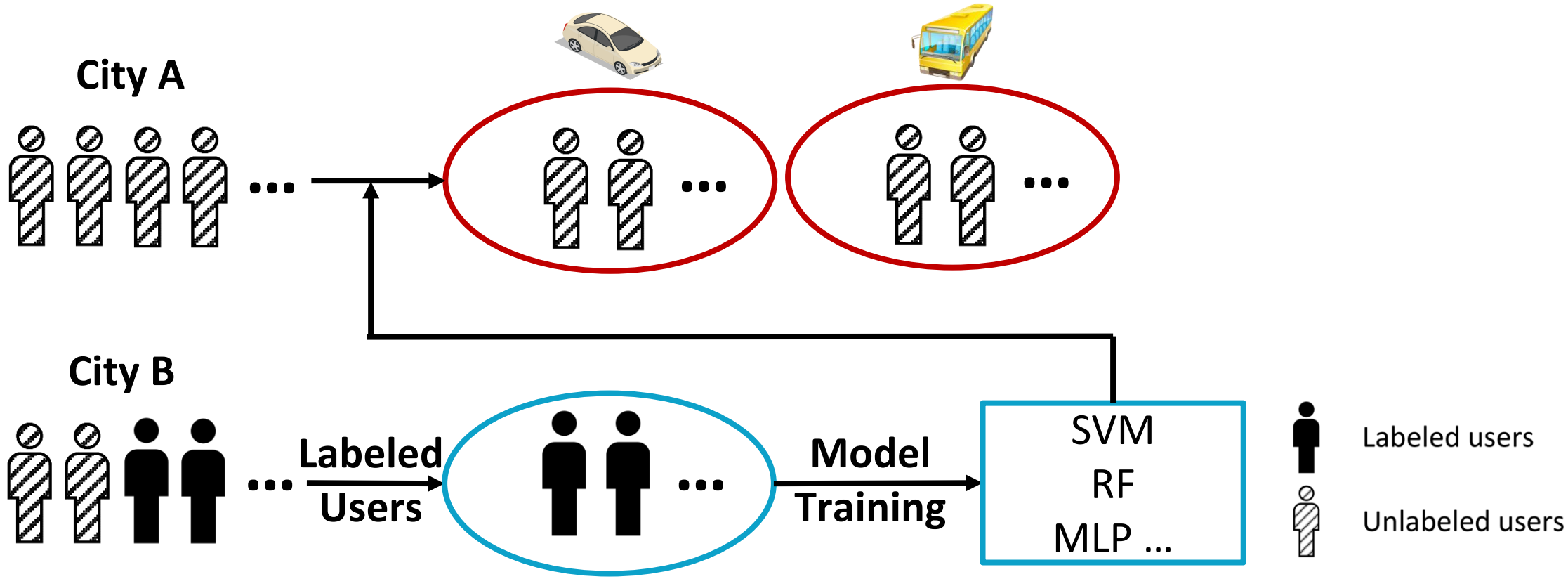


Clustering



Mode Inference Model

Scenario 2: Without Labeled Users:



Evaluation

Groundtruth:

- ws/mapapi/navigation/bus/ext

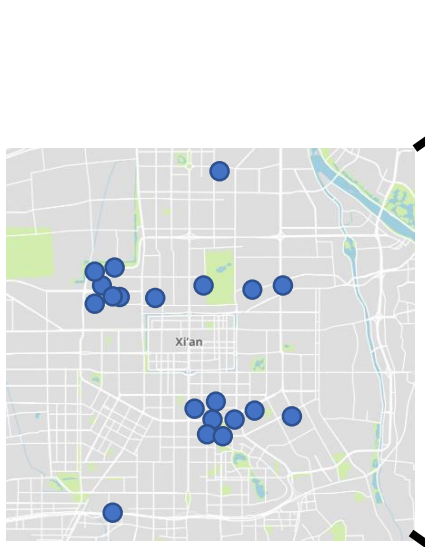
Shenyang

Transportation mode	# Groundtruth users
Car	679
Public transportation	633

Dalian

Transportation mode	# Groundtruth users
Car	813
Public transportation	464

Evaluation: Scenario 1



MFR Data

Mobility Features

Trips

Previous Methods

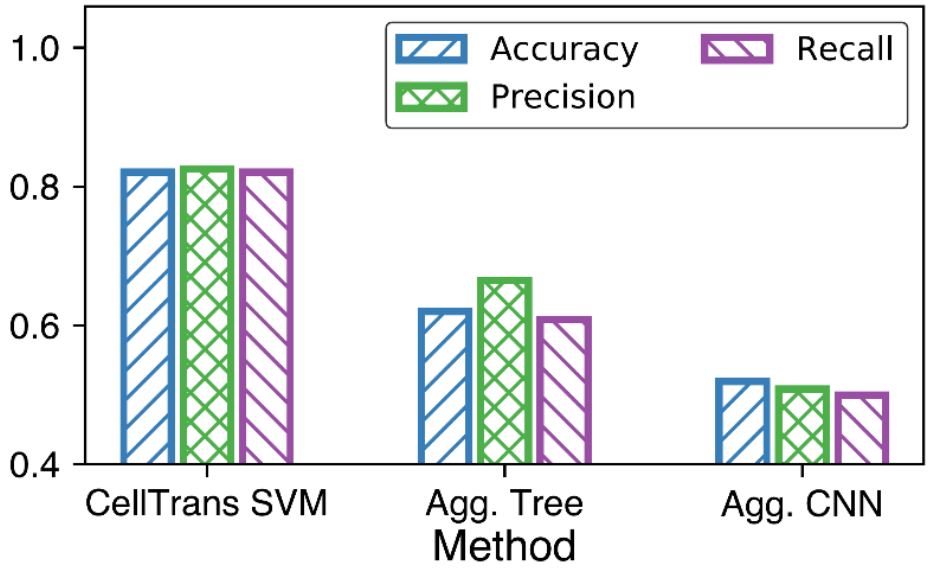
Aggregate

SVM

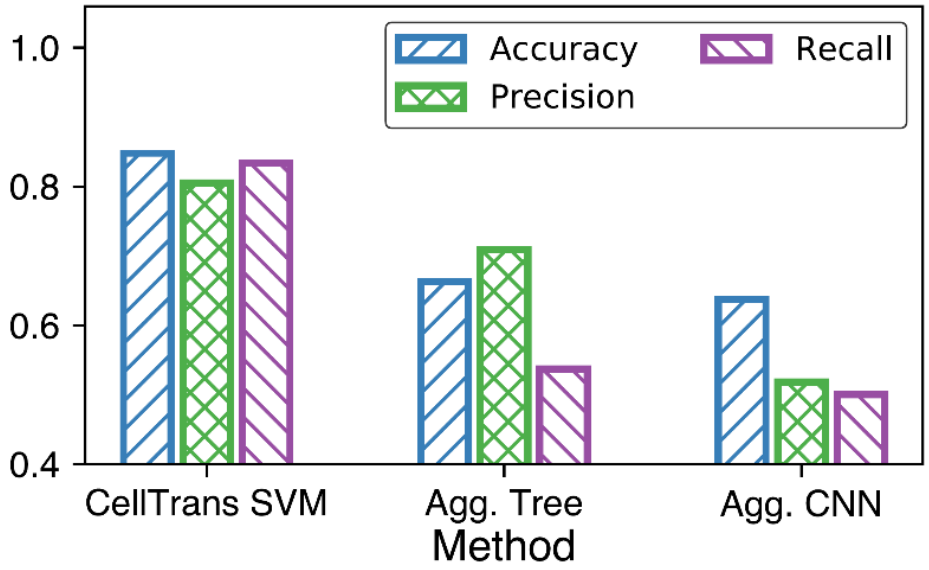
Main Mode

Evaluation: Scenario 1

- In Shenyang, CellTrans improves the accuracy by 20%.
- In Dalian, CellTrans improves the accuracy by 19%.



Shenyang



Dalian

Evaluation: Scenario 1

- Evaluate the trained model at urban scale.

Evaluation: Scenario 1

Distribution of car/public transportation users' homes:

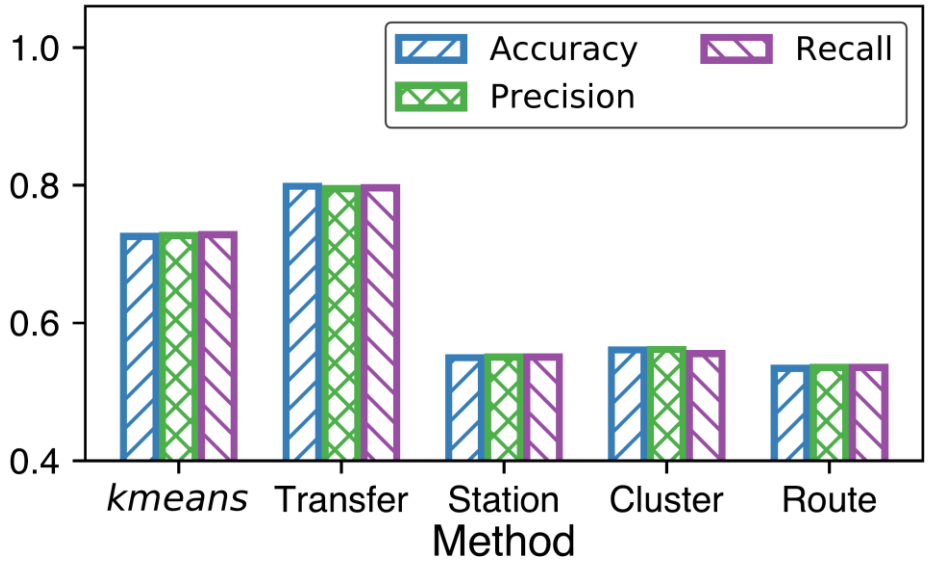
- A: High-end residential areas -> More car users.
- B: Universities -> More public transportation users.

Shenyang, car users

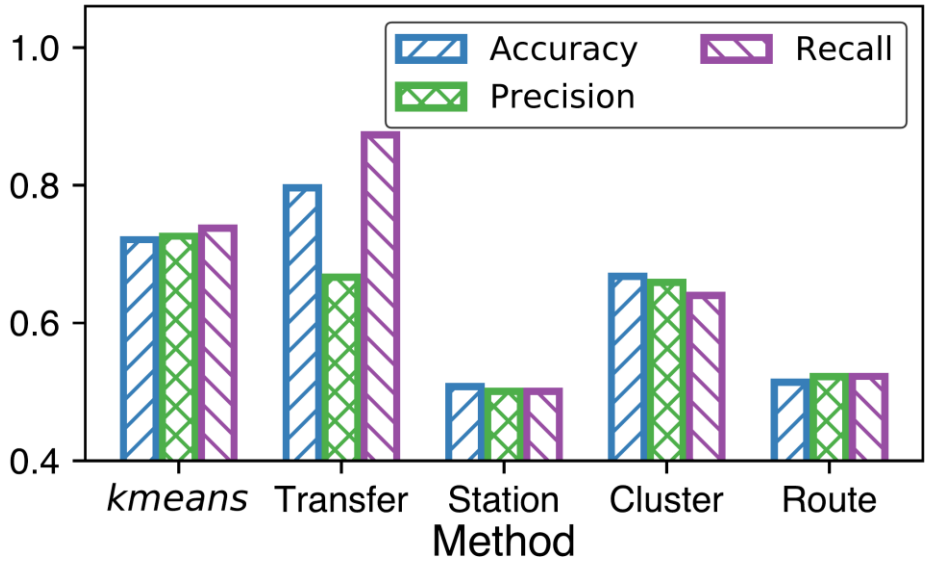
Shenyang, public transportation users

Evaluation: Scenario 2

- Our methods outperform previous methods in both cities.
- The transferred model achieves the best results.



Shenyang

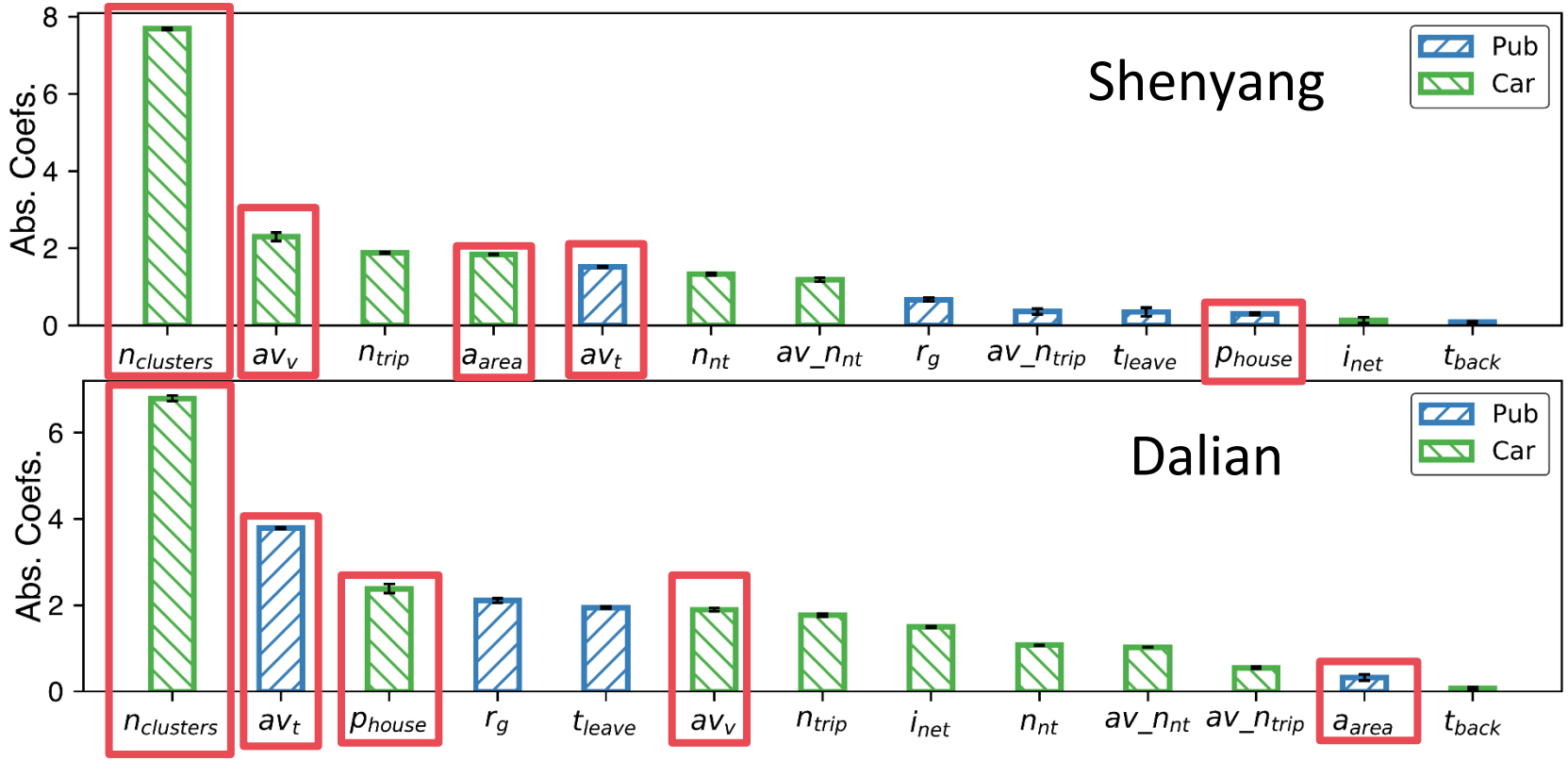


Dalian

Evaluation: Feature Importance

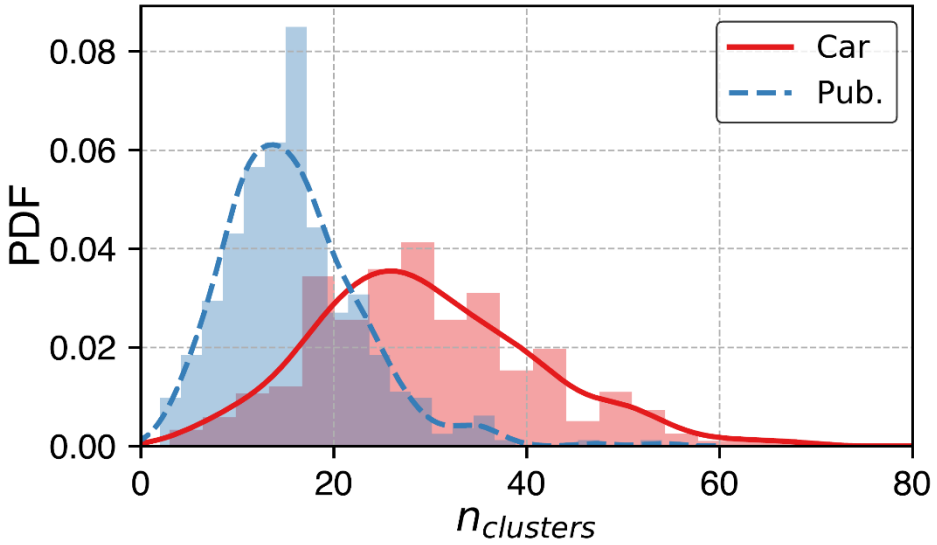
How important is each feature? -> The coefficients in Linear SVM.

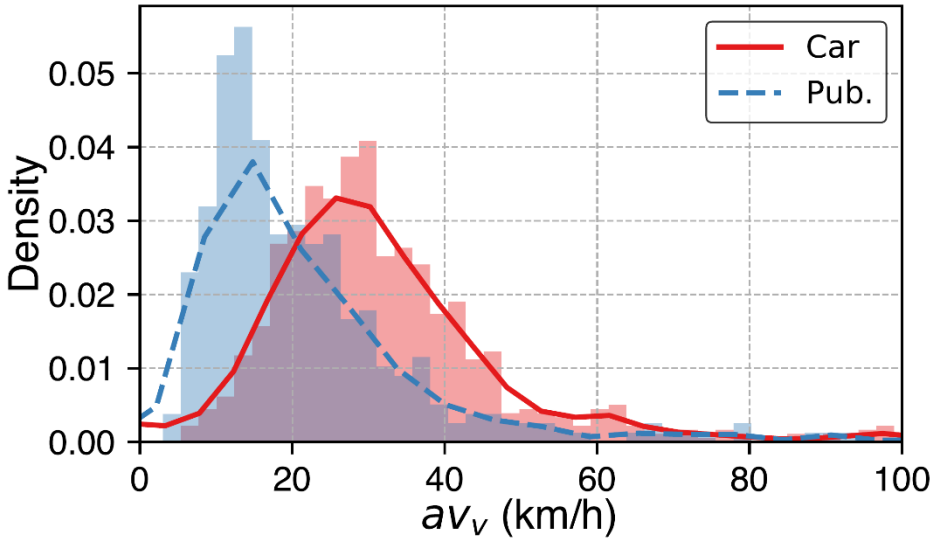
- Some features are important in both cities.



Evaluation: Feature Distribution

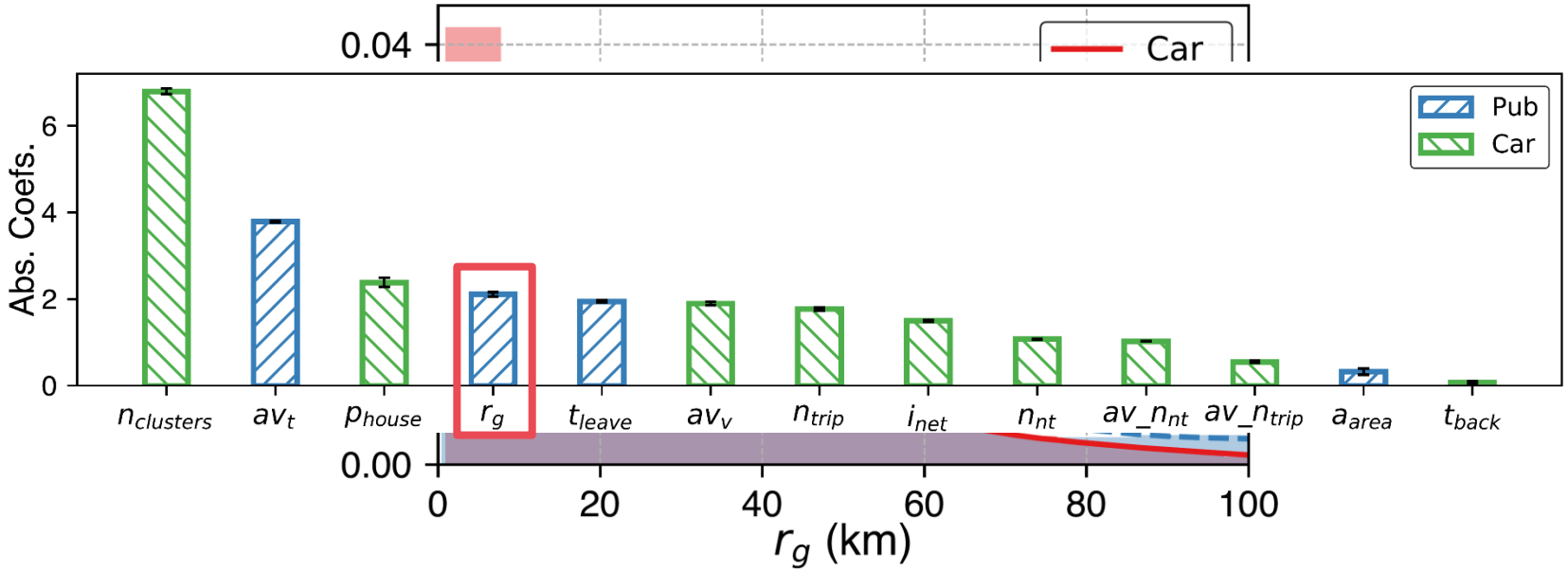
- Some features have obviously different distribution between two modes.





Evaluation: Feature Distribution

- Some features have similar distribution, but they are still helpful to differentiate main transportation modes.



Summary

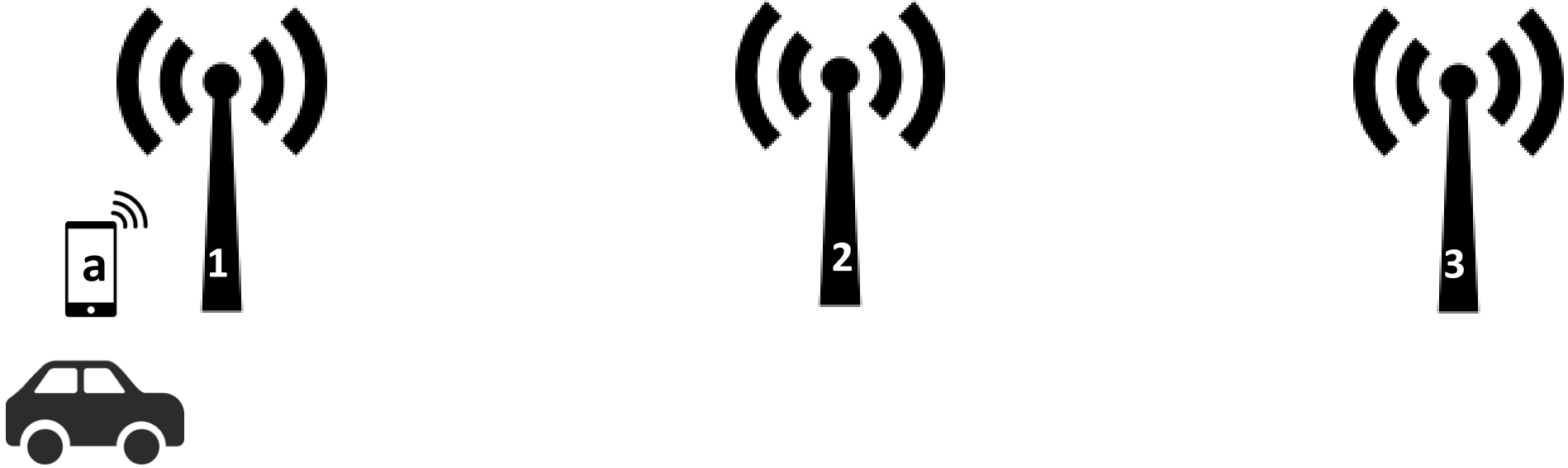
- We present CellTrans, a novel framework to survey users' main transportation modes (public transportation or private car) at urban scale.
- We devise techniques to extract various mobility features from noisy cellular data that are pertinent to users' transportation modes.
- We carry out comprehensive experiments to evaluate the performance of CellTrans on two large-scale cellular datasets.

Thanks!

Q&A

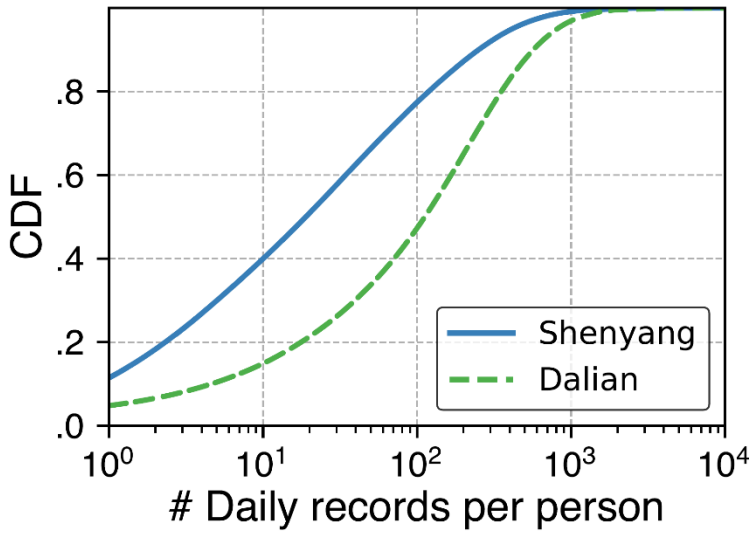
Dataset

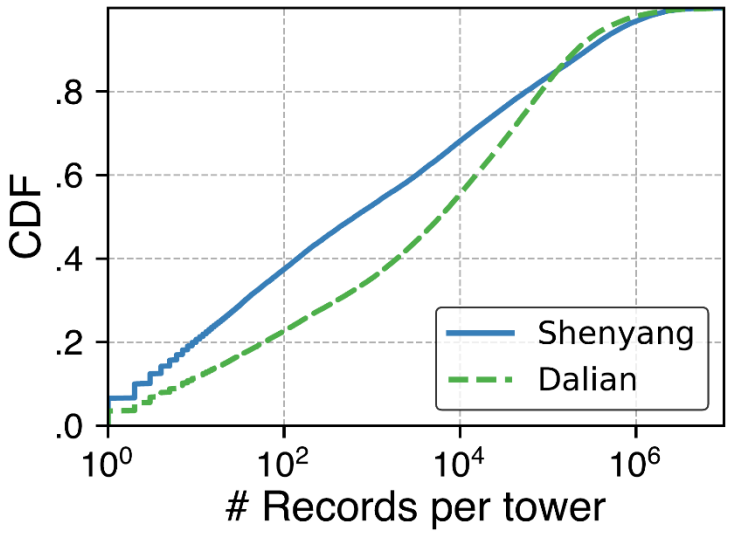
User	Time	Tower	Tower location	HTTP host	HTTP URI

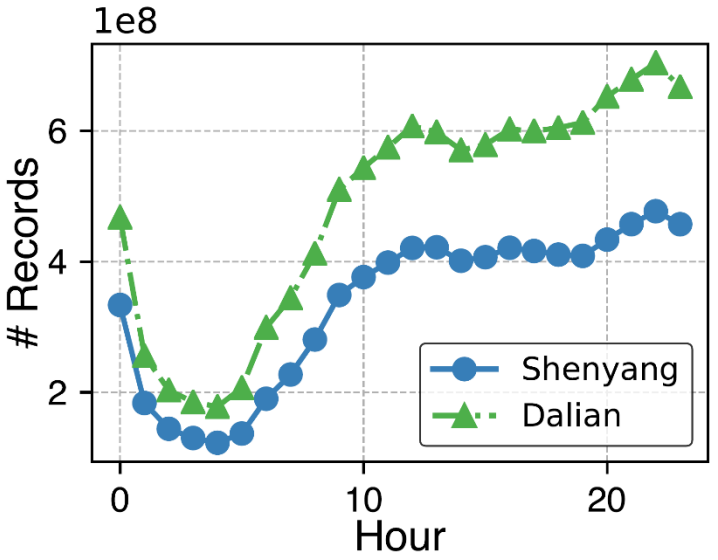


Dataset

The distribution of cellular data is uneven.



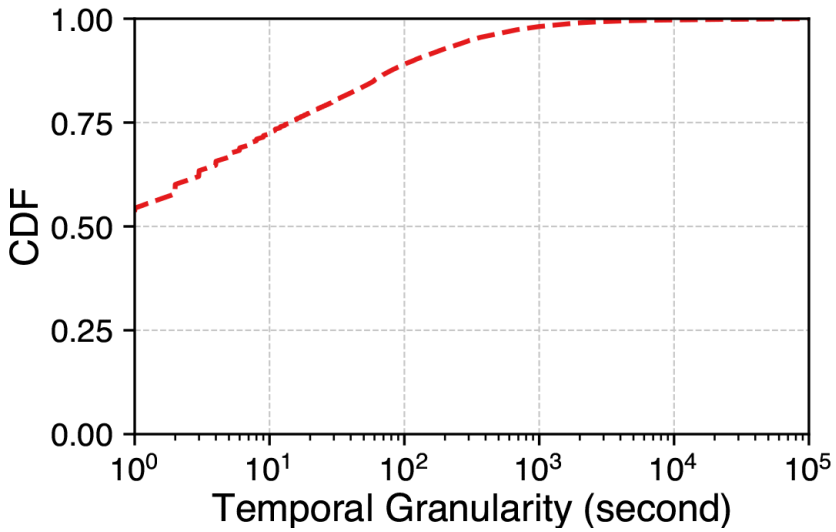




Preprocessing

The preprocessing module deals with two problems of cellular data:

Oscillation[1]



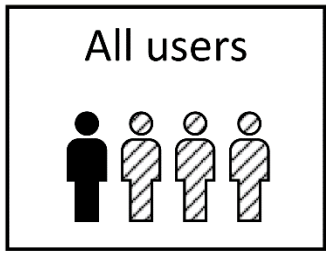
Bursty Sampling[2]

[1] Ling Qi, Yuanyuan Qiao, Fehmi Ben Abdesslem, Zhanyu Ma, and Jie Yang. 2016. Oscillation Resolution for Massive Cell Phone Traffic Data. MobiData '16

[2] Yi Zhao, Zimu Zhou, Xu Wang, Tongtong Liu, Yunhao Liu, and Zheng Yang. 2019. CellTradeMap: Delineating Trade Areas for Urban Commercial Districts with Cellular Networks. INFOCOM 2019.

Mode Inference Model

Scenario 1: With Labeled Users:

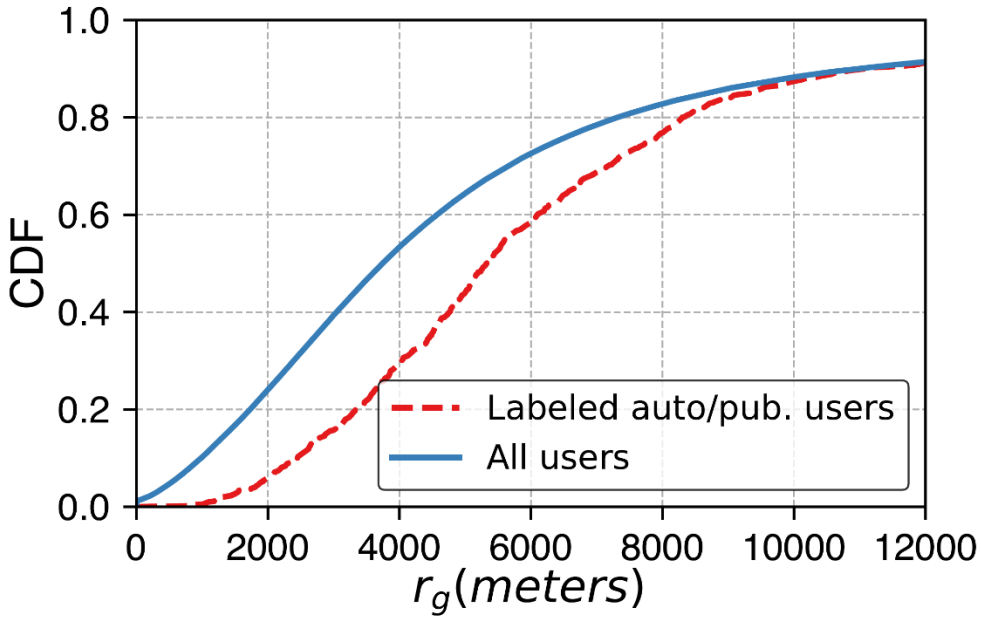


Labeled users

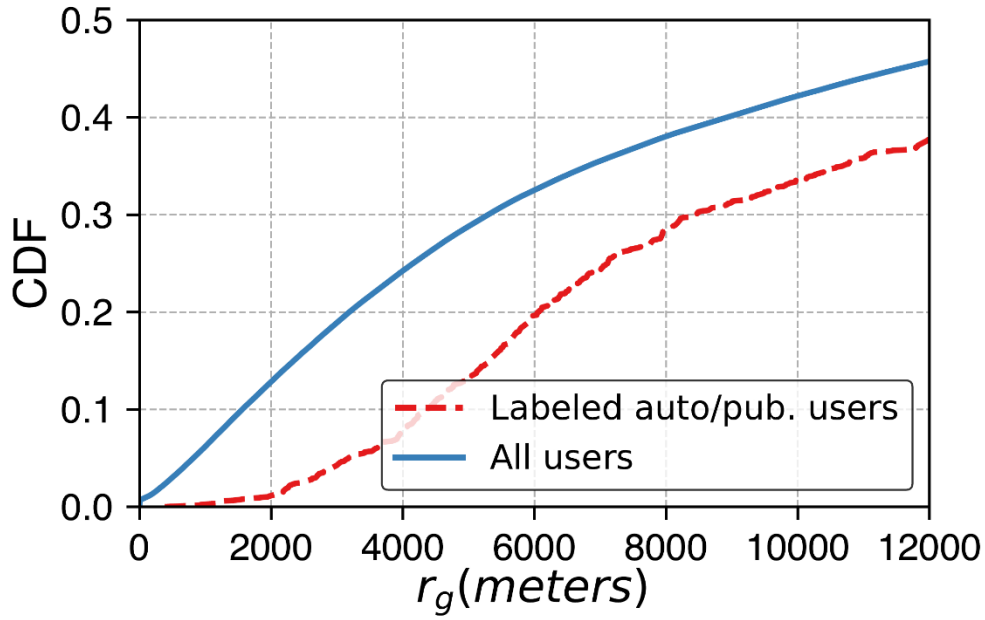
Unlabeled users

Value of Rg

CDF of rg for all users and car/pub. users.



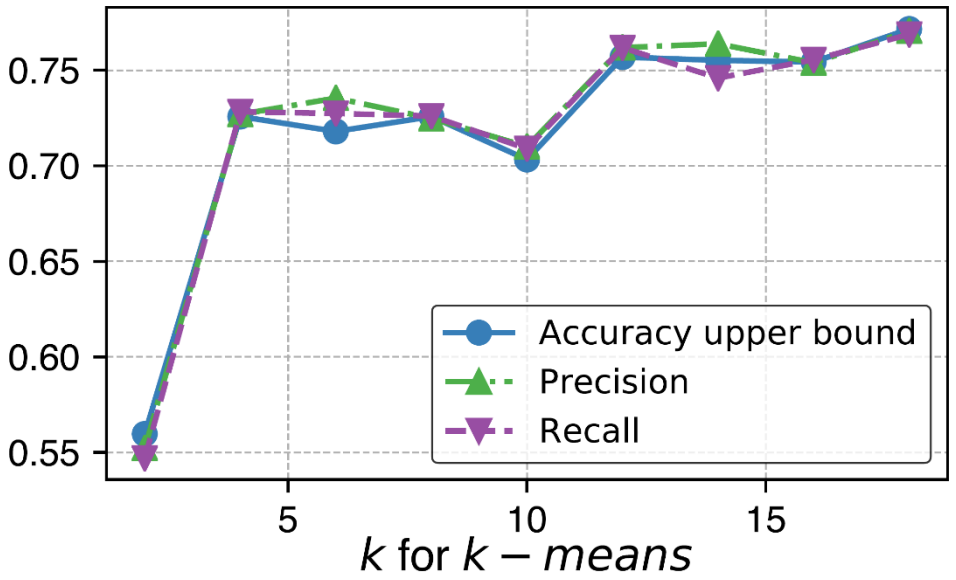
Shenyang



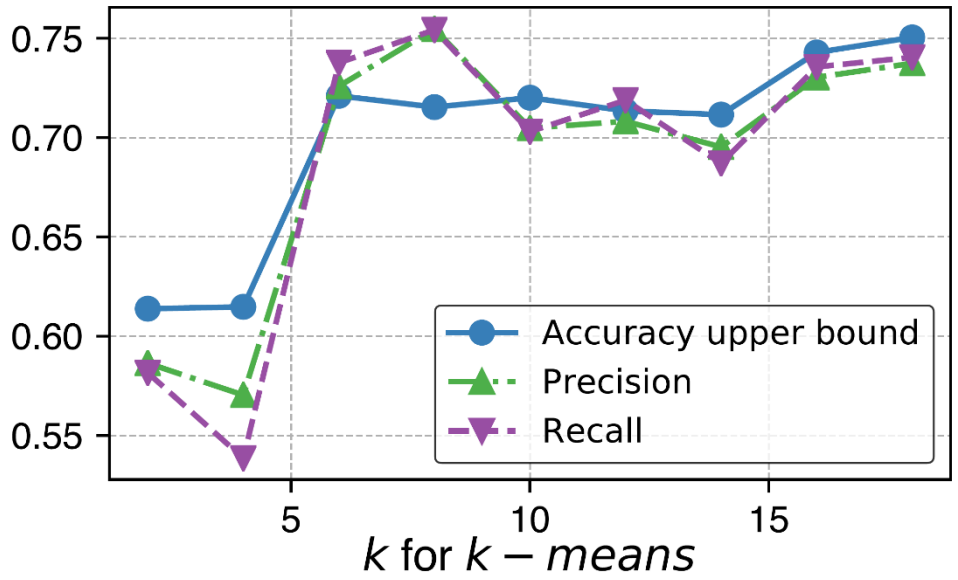
Dalian

Selection of k in K-means

Accuracy with k.

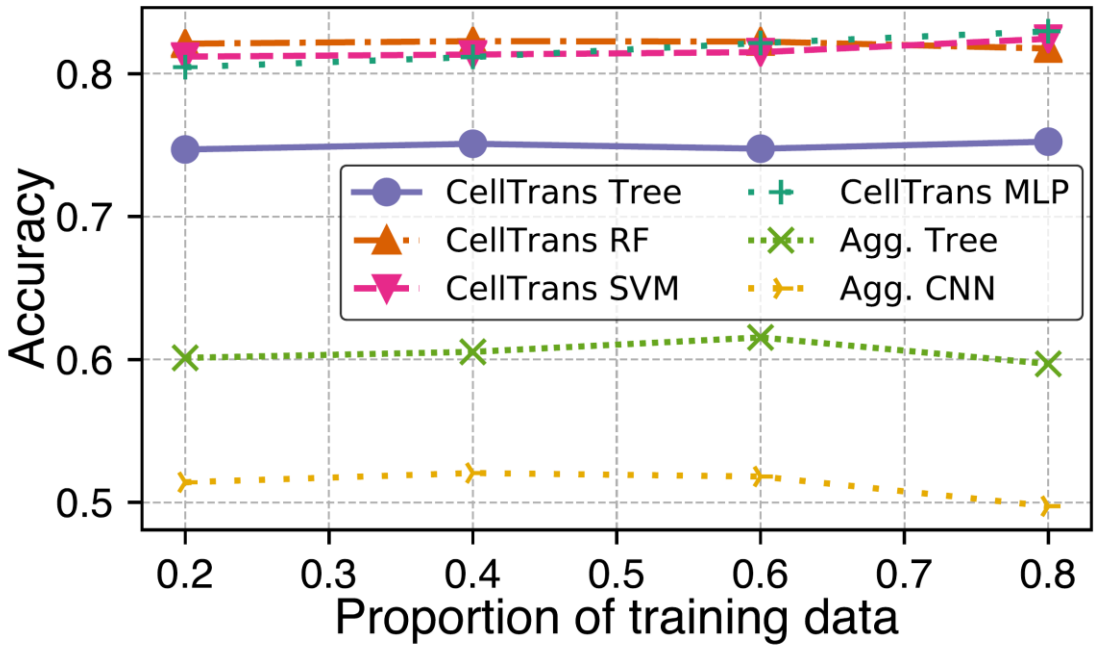


Shenyang

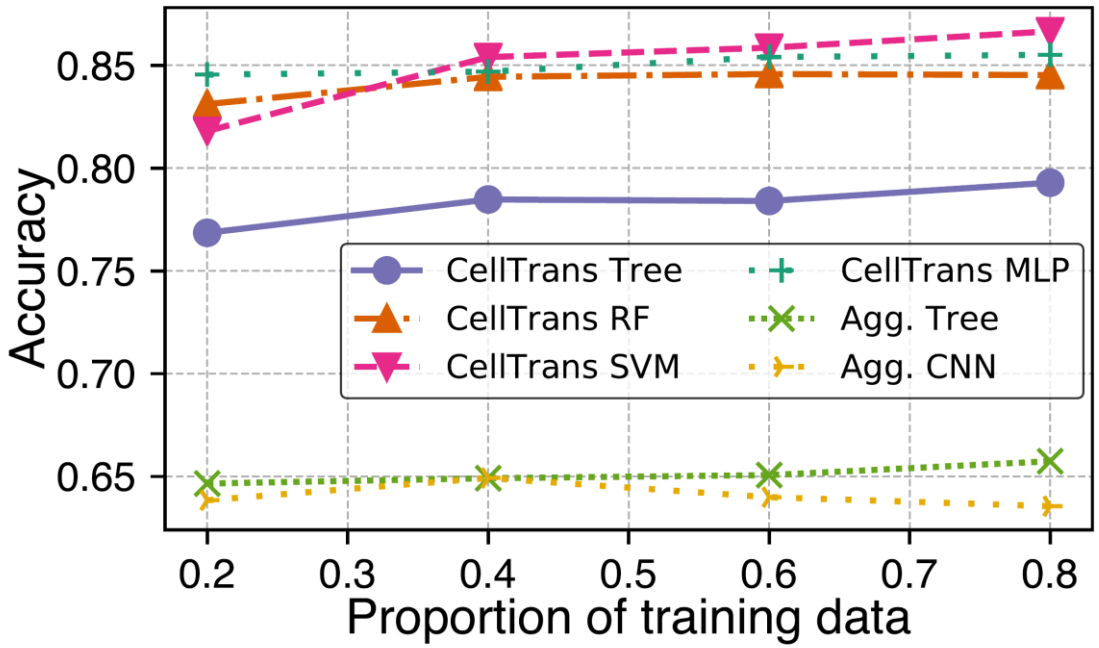


Dalian

How many labeled users do we need?



Shenyang



Dalian