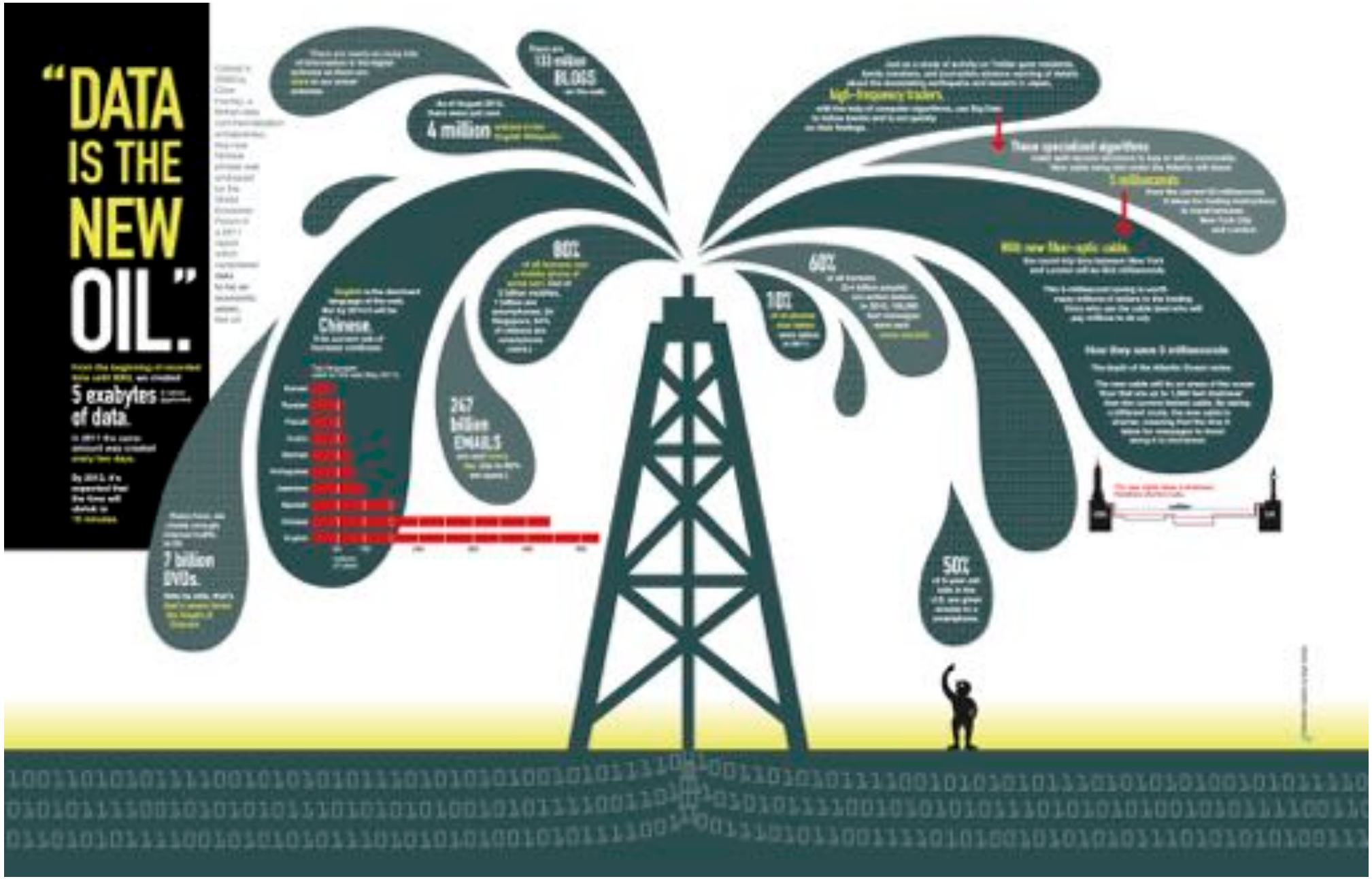
Urban Data Scarcity: from Sparse Crowdsensing to Urban Transfer Learning

Leye Wang, Peking University

leyewang@pku.edu.cn, https://sites.google.com/site/wangleye/

2019.08.17


"DATA IS THE NEW OIL."
5 exabytes of data.
7 billion DVDs.
Chinese.
4 million
80%
247 billion EMAILS
60%
10%
50%
High-frequency traders.
These specialized algorithms
5 milliseconds
How they save 5 milliseconds

But data is not always available

Urban Data Scarcity

Collect New Data

Exploit Existing Data

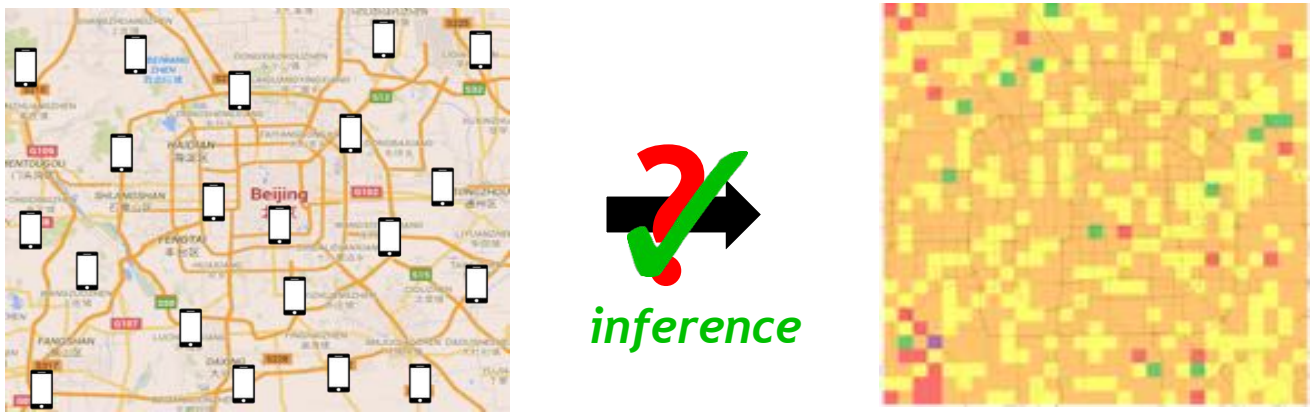
Sparse Crowdsensing

Urban Transfer Learning

Smart City

Sparse Crowdsensing

Sparsely Sensing → Full Map

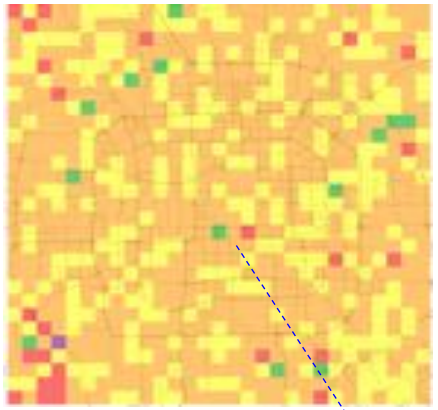


Sense Less, Infer More

	Traditional MCS	Sparse MCS
High Quality	high/full coverage	high inference accuracy (probably sparse coverage)

Sparse mobile crowdsensing: challenges and opportunities. *IEEE Communications Magazine (2016).*

Key Issues in Sparse MCS



CELL SELECTION

DATA INFERENCE

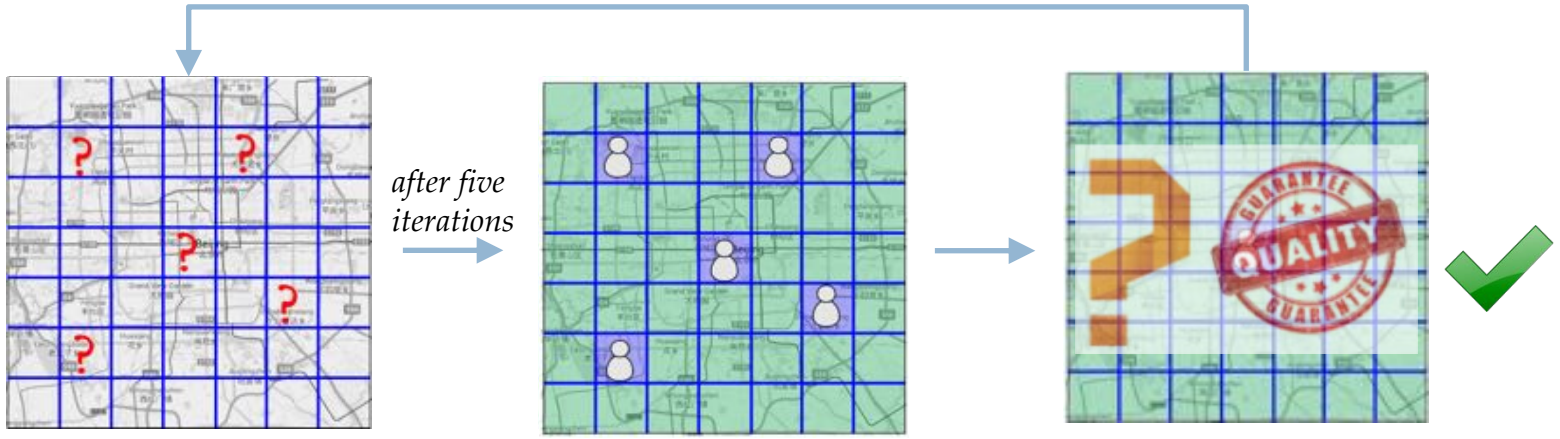
QUALITY ASSESSMENT

Three-step framework

Area: fixed-size cells (e.g.100m*100m); *Duration*: fixed-length cycl

Step 3. Quality Assessment

Use collected data to estimate inference error of unsensed cells.



Step 1: Cell Selection *Step 2: Data Inference* *Step 3: Quality Assessment*

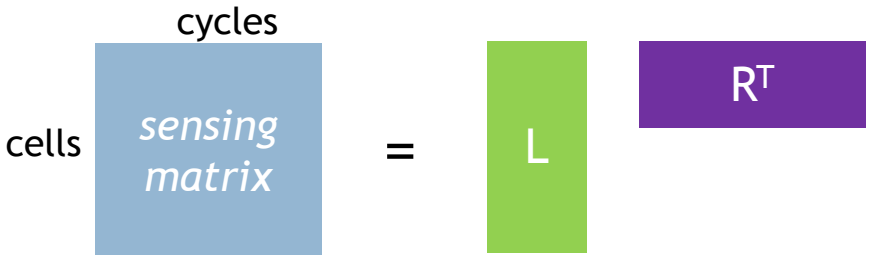
CCS-TA: quality-guaranteed online task allocation in compressive crowdsensing. *UbiComp'15*.

Data Inference

- Feature of Environment Data
  - ***Spatial and temporal*** *correlations*
  - ***Low-rank*** *property in sensing matrix*

sensing matrix: entry [*i*, *j*] means the sensing data of cell *i* in cycle *j*.

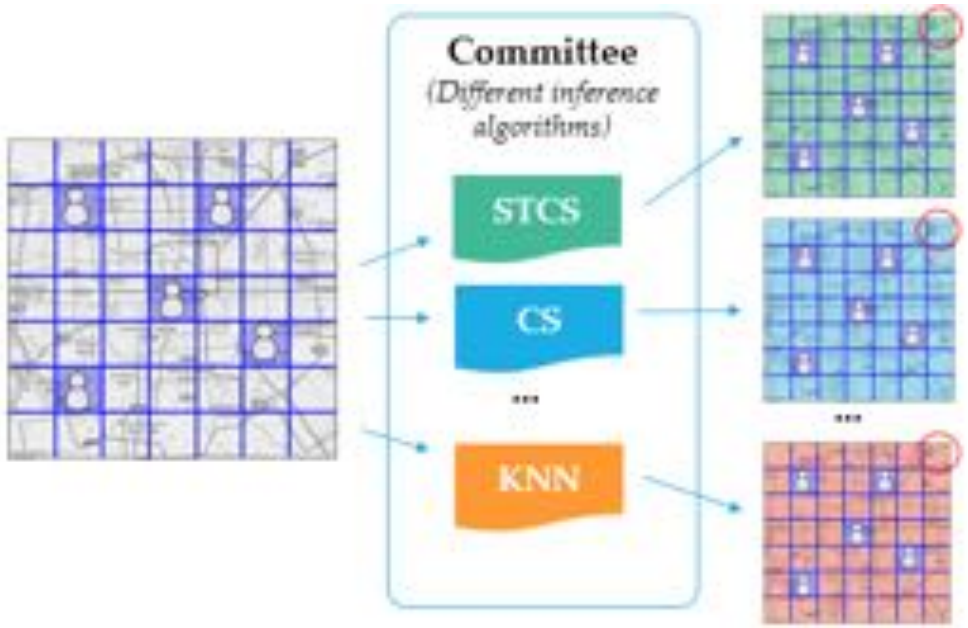
- ***Spatio-Temporal Compressive Sensing (STCS)*** *considers* ***spatial and temporal correlations****, and* ***low-rank*** *property all together.*



$$\min \lambda_r(||L||_F^2 + ||R||_F^2) + ||LR^T \circ S - C||_F^2 + \lambda_s||\mathbb{S}(LR^T)||_F^2 + \lambda_t||(LR^T)\mathbb{T}^T||_F^2$$

Cell Selection

- Allocate Task to the cell that is hard to be inferred. Then which cell?
  - *Intuition*: If different inference algorithms get significantly different inferred values for a cell, then this cell is said to be hard to be inferred.
- Called ***Query by Committee (QBC)***



Selected Cell:
Its ... vary most significantly

Quality Assessment

- ***LOO-BI***
  - ***Step 1*: Leave-one-out Resampling**
  - ***Step 2*: Bayesian Inference**
    - *probability distribution* of error

Example

***Leave-one-out Result*:**
One cell out of the five (1/5) is inferred with wrong air quality category (good/medium/bad)

Bayesian Inference:
Classification error ~ *Beta(2,5)*

CDF of Beta(2, 5)

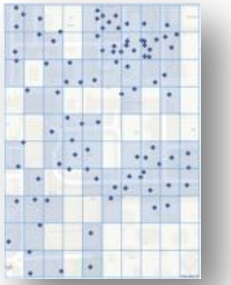
P (classification error ≤ 0.5)=0.9

***Current quality**: In 90% of cycles, the classification error is ≤ 50%*

Experiment setting

- Datasets
  - **Temperature**: 57 cells (30m*50m), 30-min cycle; mean absolute error
  - **PM2.5**: 36 cells (1km*1km), 1-hour cycle; classification error (six AQI categories)
- Our Method
  - **CCS-TA**
- Baselines
  - **RAND-TA**: randomly selecting next sensing cell.
  - **FIX-TA-k**: keep same number of sensing cells (k) in each cycle.

temperature

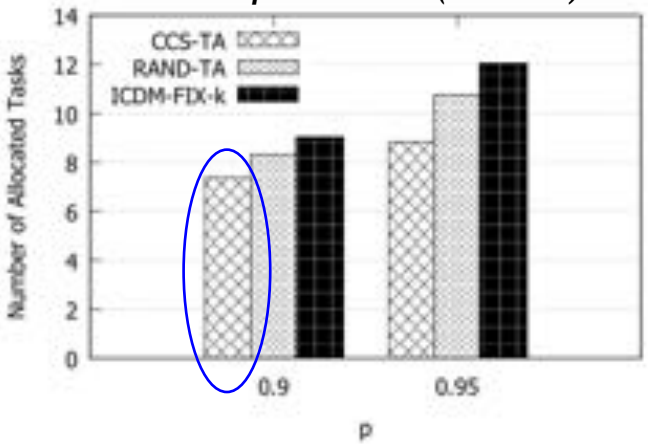


PM2.5

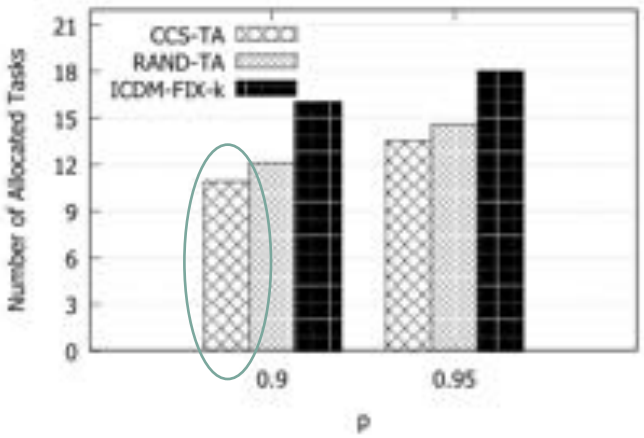
Q. How many sensed cells are necessary for sparse crowdsensing and baselines for the same quality requirement?

Number of Cells Required for CCS-TA

Temperature (0.25°C)



PM2.5 (25%)



Sense 13% of cells →
Mean absolute error of temperature ≤ 0.25°C in 90% of cycles

Sensing 29% of cells →
Classification error of PM2.5 ≤ 25% in 90% of cycles

Initial Proposal of the idea of Sparse Crowdsensing:

- **L. Wang**, D. Zhang, A. Pathak, C. Chen, H. Xiong, D. Yang, Y. Wang. CCS-TA: Quality-Guaranteed Online Task Allocation in Compressive Crowdsensing. UbiComp 2015
- **L. Wang**, D. Zhang, Y. Wang, C. Chen, X. Han, A. Mhamed. Sparse Mobile Crowdsensing: Challenges and Opportunities. IEEE Communications Magazine, vol. 54, no. 7, pp. 161-167, 2016

From Single Task to Multi Tasks:

- **L. Wang**, D. Zhang, D. Yang, A. Pathak, C. Chen, X. Han, H. Xiong, Y. Wang. SPACE-TA: Cost-Effective Task Allocation Exploiting Intradata and Interdata Correlations in Sparse Crowdsensing. ACM Transactions on Intelligent Systems and Technology, vol. 9, no. 2, pp. 20:1-20:28, 2018

Privacy Protection:

- **L. Wang**, D. Zhang, D. Yang, B. Y. Lim, X. Ma. Differential Location Privacy in Sparse Mobile Crowdsensing. ICDM 2016: 1257-1262.
- T. Zhou, Z. Cai, B. Xiao, **L. Wang**, M. Xu, Y. Chen. Location Privacy-Preserving Data Recovery for Mobile Crowdsensing. UbiComp 2018.

Reinforcement Learning for Cell Selection:

- **L. Wang**, W. Liu, D. Zhang, Y. Wang, E. Wang, Y. Yang. Cell Selection with Deep Reinforcement Learning in Sparse Mobile Crowdsensing. ICDCS 2018
- W. Liu, Y. Yang, E. Wang, **L. Wang**, D. Zeghlache, D. Zhang. Multi-Dimensional Urban Sensing in Sparse Mobile Crowdsensing. IEEE Access vol. 7, pp. 82066-82079, 2019

Related Research from Other Teams

When Compressive Sensing Meets Mobile Crowdsensing

Authors (view affiliations)

Linghe Kong, Bowen Wang, Guihai Chen

Book

Cost-aware compressive sensing for networked sensing systems

2015, Information Processing in Sensor Networks, pp 130–141

Liwen Xu (Tsinghua University), Xiaohong Hao (Tsinghua University), Nicholas D. Lane (Microsoft), Xin Liu (University of California, Davis), Thomas Moscibroda (Microsoft)

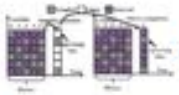
Density-aware compressive crowdsensing

2017, Information Processing in Sensor Networks, pp 29–39

Xiaohong Hao (Tsinghua University), Liwen Xu (Tsinghua University), Nicholas D. Lane (University College London), Xin Liu (University of California, Davis), Thomas Moscibroda (Microsoft)

Active Sparse Mobile Crowd Sensing Based on Matrix Completion

Full Text: PDF Get this Article



2019 Article

Authors:

Kun Xie	Hunan University, Changsha, China
Xiaocan Li	Hunan University, Changsha, China
Xin Wang	Stony Brook University, Stony Brook, NY, USA
Gaogang Xie	Institute of Computing Technology & University of Chinese Academy of Sciences, Beijing, China
Jigang Wen	Institute of Computing Technology & Chinese Academy of Sciences, Beijing, China
Dafang Zhang	Hunan University, Changsha, China

Bibliometrics

- Citation Count: 0
- Downloads (cumulative): 107
- Downloads (12 Months): 107
- Downloads (6 Weeks): 61

Published in:

- Proceeding

SIGMOD '19 Proceedings of the 2019 International Conference on Management of Data
Pages 195-210

Amsterdam, Netherlands — June 30 - July 05, 2019
ACM New York, NY, USA ©2019
table of contents ISBN: 978-1-4503-5643-5 doi>10.1145/3299869.3319856

More with less: lowering user burden in mobile crowdsourcing through compressive sensing

2015, Ubiquitous Computing, pp 659–670

Liwen Xu (Tsinghua University), Xiaohong Hao (Tsinghua University), Nicholas D. Lane (Bell Labs), Xin Liu (University of California, Davis), Thomas Moscibroda (Microsoft)

Urban Data Scarcity

Collect New Data

Exploit Existing Data

Sparse Crowdsensing

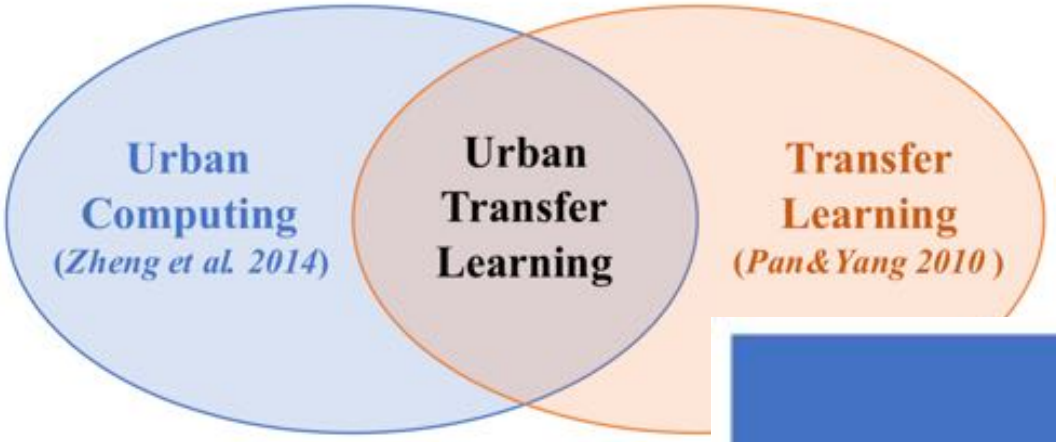
Urban Transfer Learning

Smart City

Transfer Learning

Instance Transfer, Feature Transfer, Model Transfer …

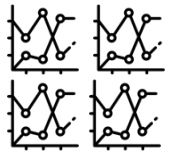
Urban Transfer Learning



	Transfer Learning on Recommendation [5]	Urban Transfer Learning on Traffic Prediction [17]
Source Domain	Movie Rating	City 1's Traffic
Target Domain	Book Rating	City 2's Traffic
Remark	**Single data modality** (*user-item ratings*)	1. **Heterogeneous data modalities** (*daytime*, *weather*, *check-ins*) 2. **Spatio-temporal patterns**

Basic Idea in Urban Transfer

Cross-Modality (heterogeneous data)

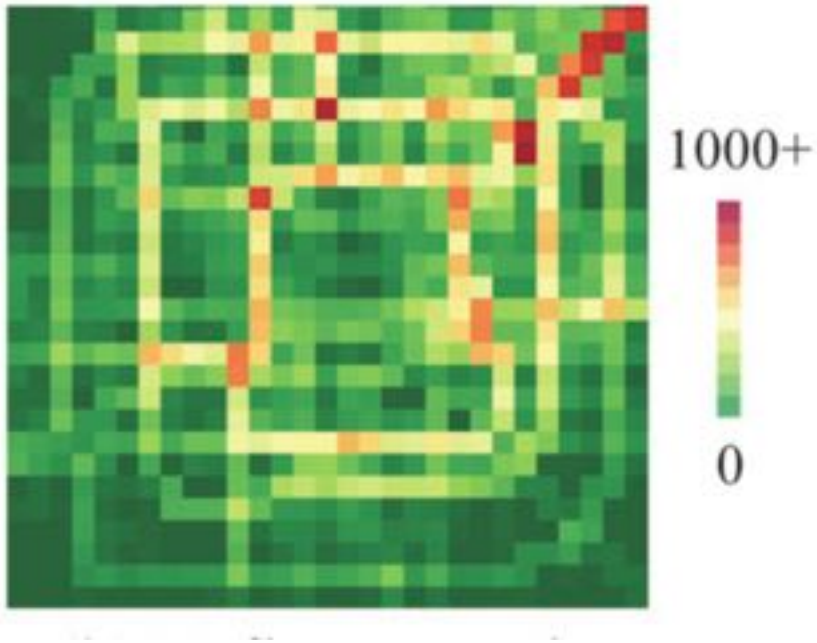


Crowd flow
(long history)

Check-ins

Cross-City (spatial transfer)

Crowd Flow Prediction



Predict inflow/outflow of a region in next 30 minutes

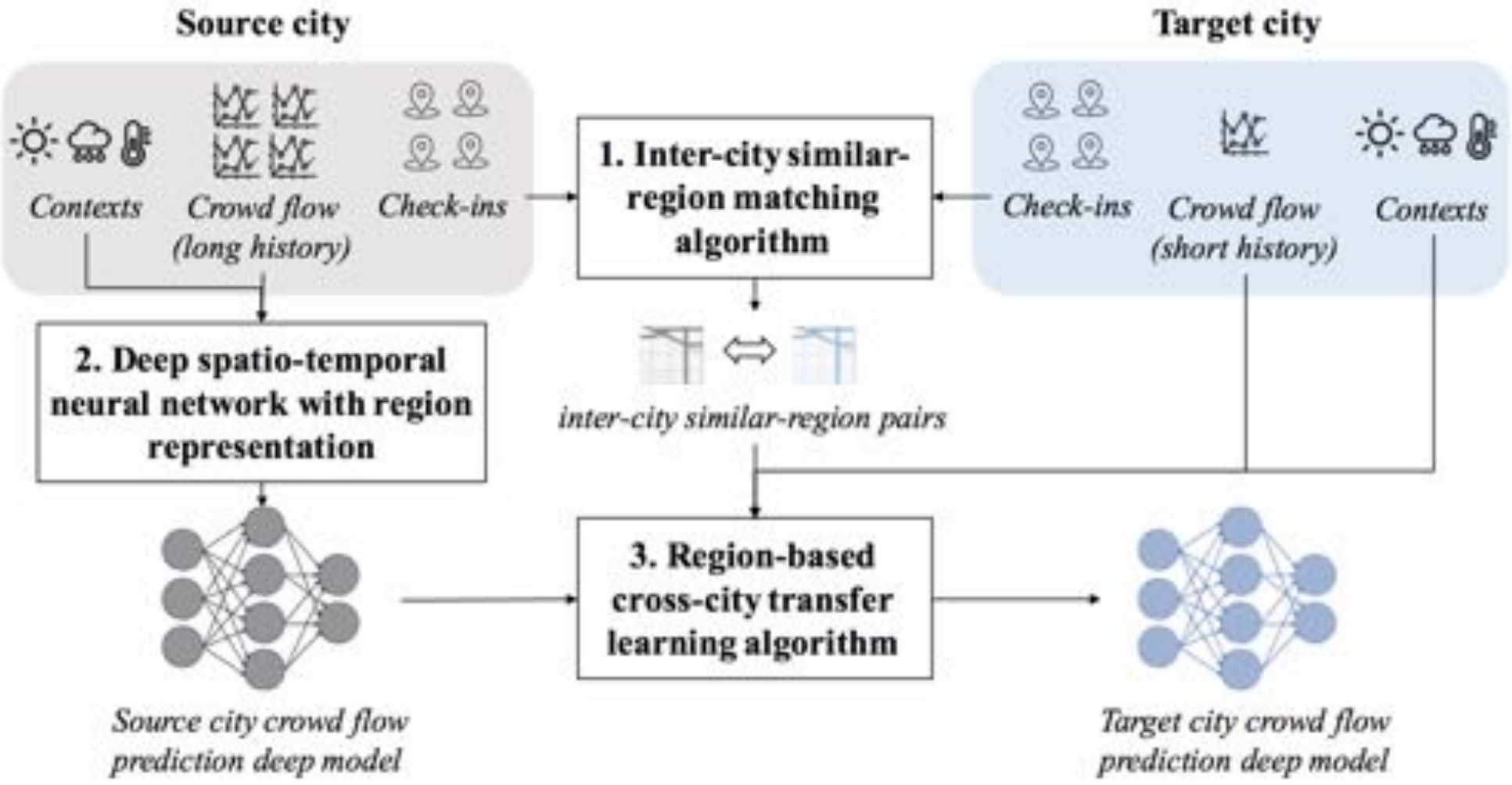
- Deep learning has shown its advantage

Zhang J. et al. "Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction." AAAI. 2017.

For a new city without so many data?

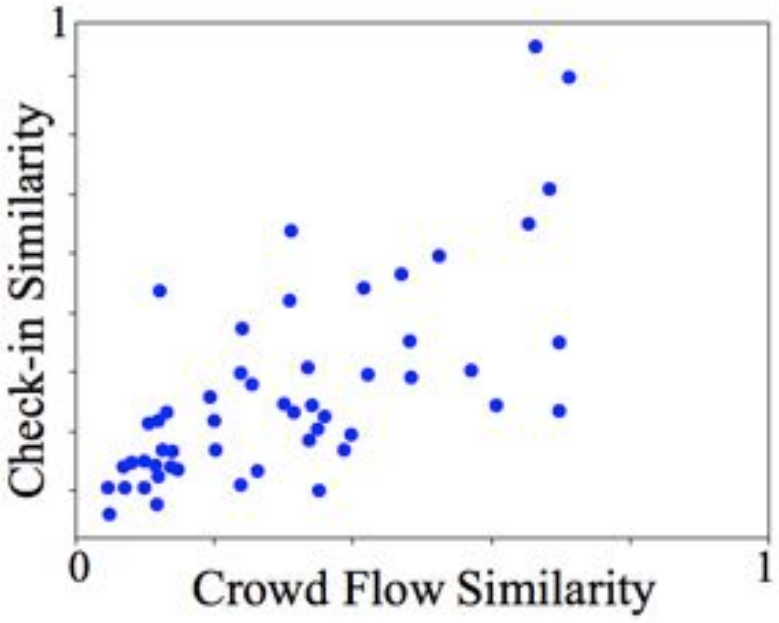
Cross-City Transfer Learning for Deep Spatio-Temporal Prediction. IJCAI, 2019.

RegionTrans

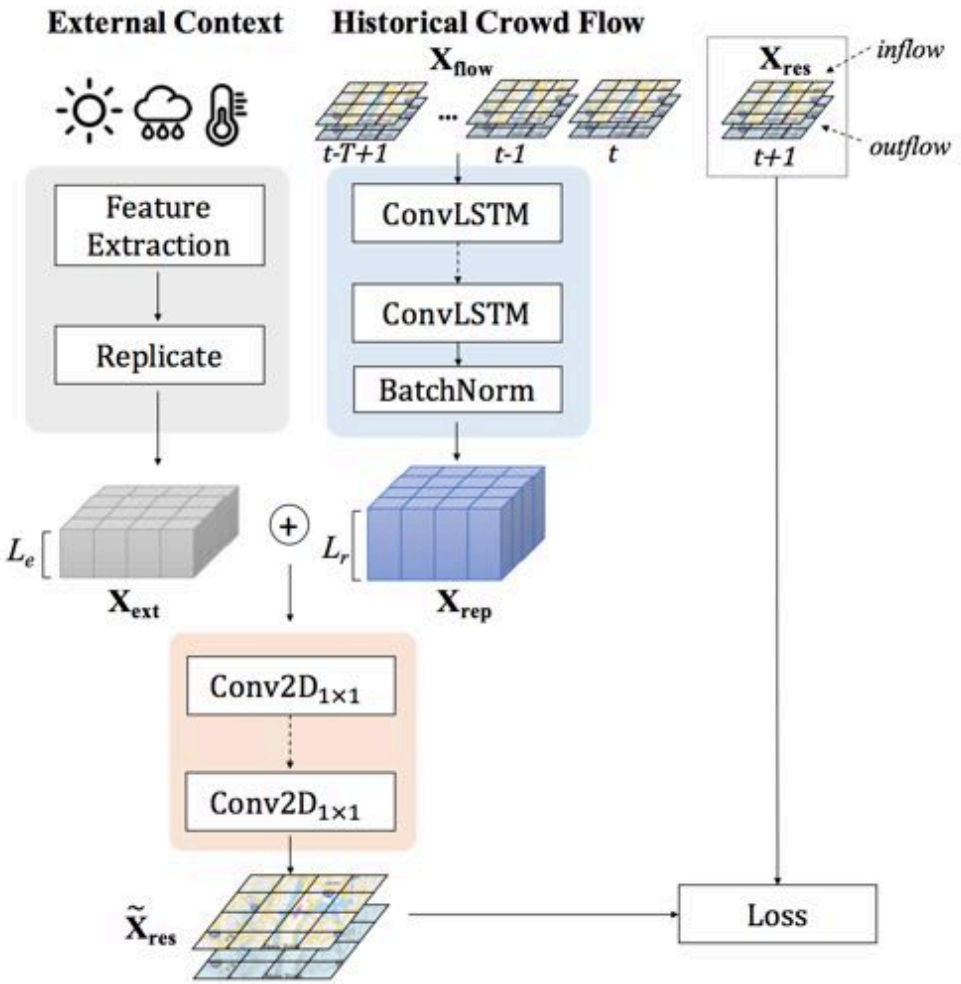


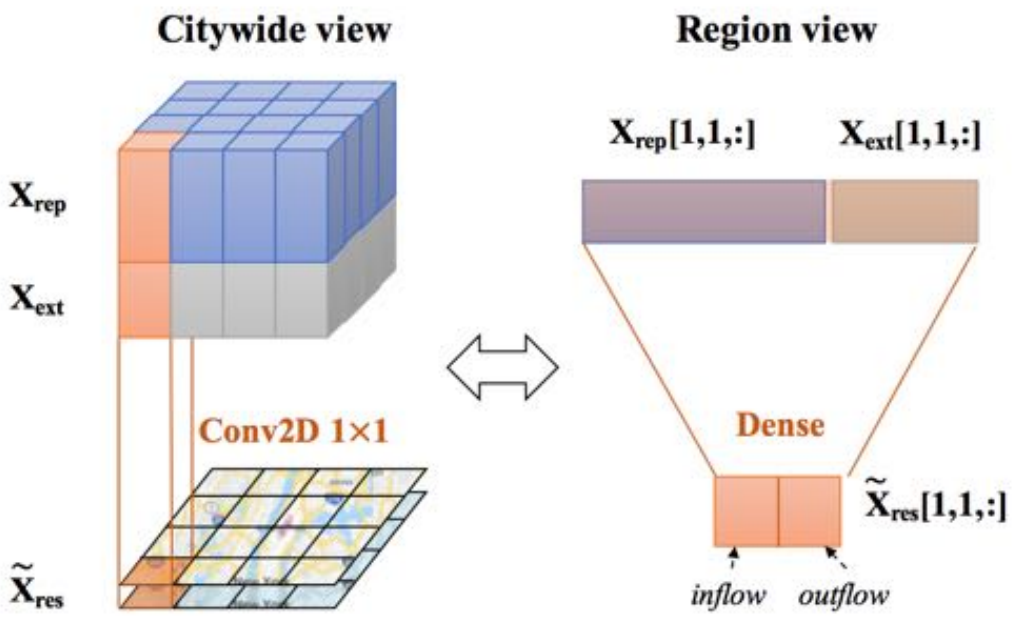
Inter-city Similar Region Matching

- Target city has very little crowd flow data
- For each target region, use **social network check-in** data to find ***top-k*** similar source regions



DNN with region representation





Cross-city Transfer Learning

- Learn a model in source city, use the parameter as the start of the target city model
- Use the little data in the target city to optimize the parameter
- Optimization objective

$$\arg\min_{\theta} \; w\left(\frac{1}{k}\sum_{1\le i\le k} ||\boldsymbol{\rho}_i \circ (\mathbf{X}_{rep} - \hat{\mathbf{X}}^i_{rep})||_2^2\right) + (1-w)||\widetilde{\mathbf{X}}_{res} - \mathbf{X}_{res}||_2^2$$

Minimize representation discrepancy

Minimize prediction error

Experiment: Bike Sharing

	Washington D.C.	Chicago
#Trip records	6,519,741	6,690,351
Time span	2015.1.1 - 2016.12.31	
Time interval	30 minutes	
Region size	$1km \times 1km$	
City map size	20×20	

	D.C.→Chicago			Chicago→D.C.		
	1-day	3-day	7-day	1-day	3-day	7-day
Target Data Only						
ARIMA	0.740	0.694	0.679	0.707	0.661	0.647
DeepST	0.771	0.711	0.636	1.075	0.767	0.691
ST-ResNet	0.914	0.703	1.053	0.869	0.738	1.054
Source & Target Data						
DeepST (FT)	0.652	0.611	0.566	0.672	0.619	0.586
ST-ResNet (FT)	0.667	0.615	0.613	0.695	0.623	0.608
RegionTrans	**0.587**	**0.576**	**0.553**	**0.600**	**0.581**	**0.573**

RegionTrans can outperform state-of-the-art deep learning with fine-tune by reducing up to 10% error (RMSE)

Robust Against Negative Transfer

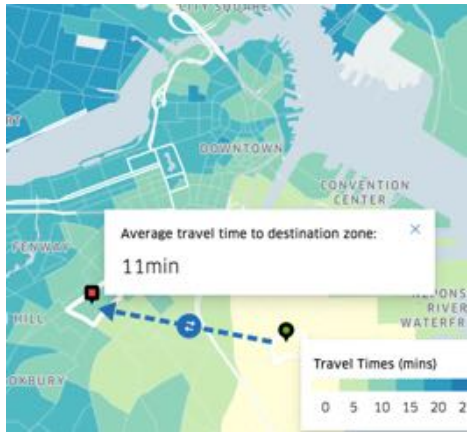
	D.C.→NYC		NYC→D.C.	
	1-day	3-day	1-day	3-day
Target Data Only				
ARIMA	0.360	0.341	0.707	0.661
DeepST	0.350	0.359	1.075	0.767
ST-ResNet	0.376	0.349	0.869	0.738
Source & Target Data				
DeepST (FT)	0.363	0.369	0.713	0.711
ST-ResNet (FT)	0.385	0.349	0.696	0.691
RegionTrans	0.328	0.305	0.665	0.593

The performance of existing deep models gets worse when using fine-tune for D.C. → NYC, but *RegionTrans* still makes an effective transfer.

Future Opportunities

Multi-Source Transfer

Privacy-Preserving Transfer



How is Uber Movement preserving the privacy of Uber riders and drivers?

Preserving rider and driver privacy is our #1 priority. All data is anonymized and aggregated to ensure no personally identifiable information or user behavior can be surfaced through the Movement tool. All data shared through Movement adheres to Uber's privacy policy, and at no point will Movement provide a means for partners to access individual driver or rider details in any way, shape or form. Further details of how we're ensuring this are available in our FAQ.

Uber Movement 共享数据平台的隐私相关声明

Smart City Development With Urban Transfer Learning

Leye Wang, Key Lab of High Confidence Software Technologies, Peking University
Bin Guo, Northwestern Polytechnical University
Qiang Yang, Hong Kong University of Science and Technology

L. Wang, B. Guo, Q. Yang, "*Smart City Development with Urban Transfer Learning*". IEEE Computer 51(12): 32-41 (2018).

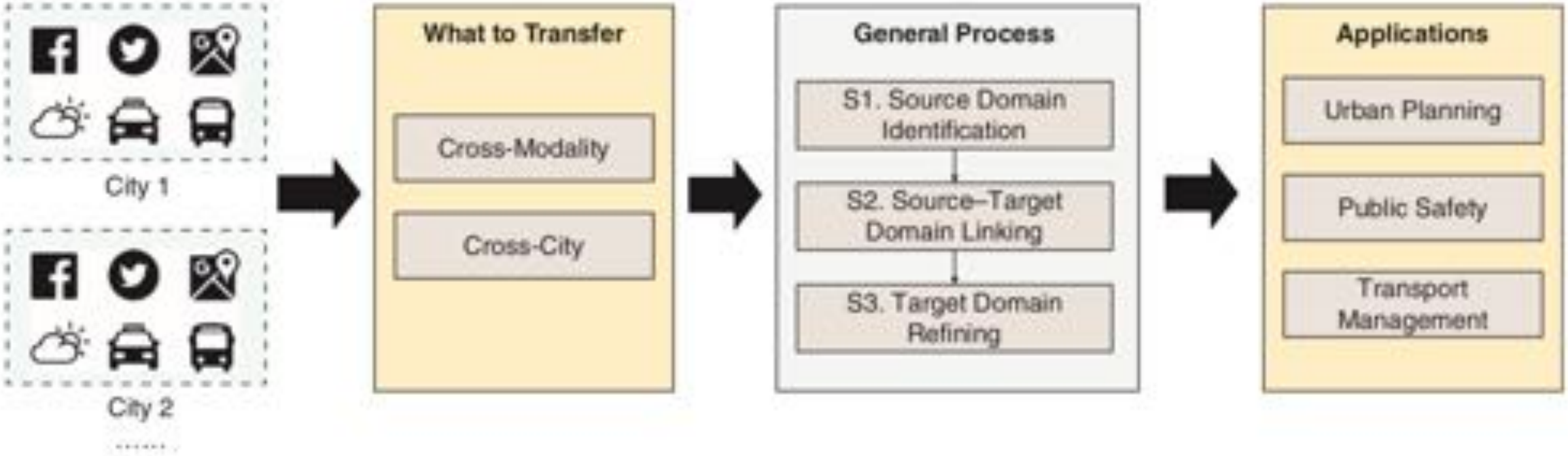
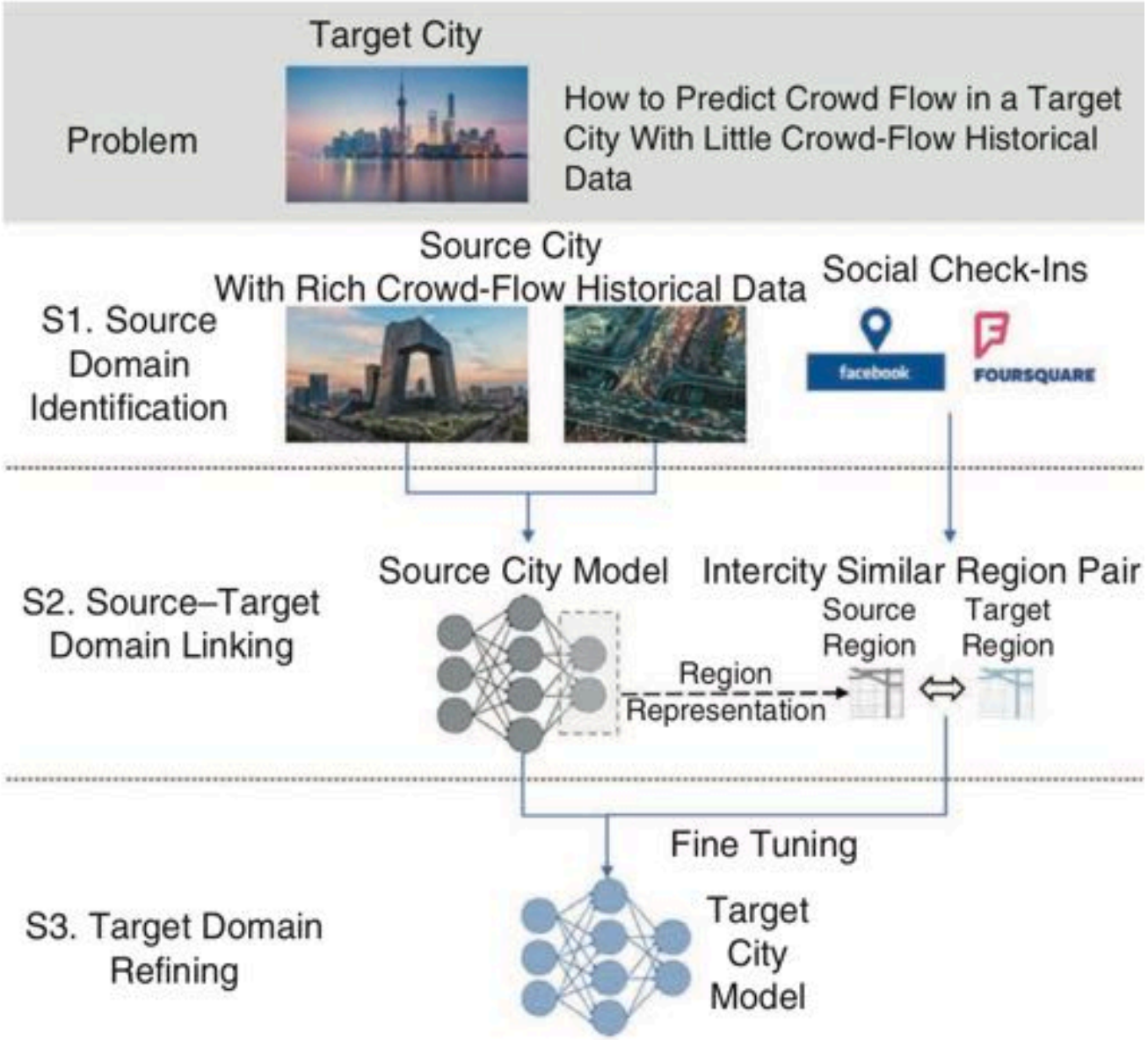
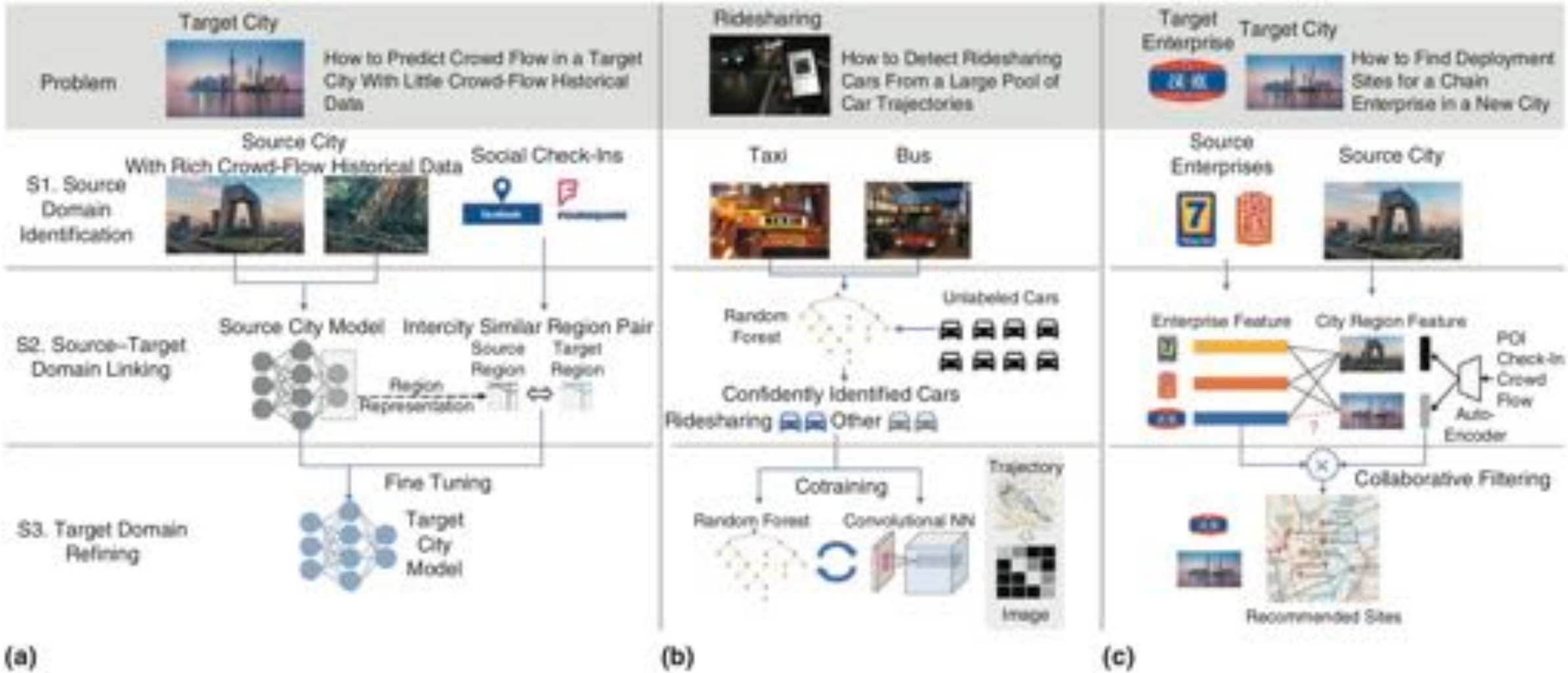


FIGURE 2. The framework of urban transfer learning.


Target City
Problem
How to Predict Crowd Flow in a Target City With Little Crowd-Flow Historical Data
Source City
With Rich Crowd-Flow Historical Data
Social Check-Ins
S1. Source Domain Identification
facebook
FOURSQUARE
Source City Model
Intercity Similar Region Pair
S2. Source–Target Domain Linking
Source Region
Target Region
Region Representation
Fine Tuning
S3. Target Domain Refining
Target City Model


Target City
Problem
How to Predict Crowd Flow in a Target City With Little Crowd-Flow Historical Data
Source City
With Rich Crowd-Flow Historical Data
Social Check-Ins
S1. Source Domain Identification
Source City Model
Intercity Similar Region Pair
Source Region
Target Region
S2. Source–Target Domain Linking
Region Representation
Fine Tuning
S3. Target Domain Refining
Target City Model
(a)
Ridesharing
How to Detect Ridesharing Cars From a Large Pool of Car Trajectories
Taxi
Bus
Random Forest
Unlabeled Cars
Confidently Identified Cars
Ridesharing
Other
Cotraining
Random Forest
Convolutional NN
Trajectory
Image
(b)
Target Enterprise
Target City
How to Find Deployment Sites for a Chain Enterprise in a New City
Source Enterprises
Source City
Enterprise Feature
City Region Feature
POI
Check-In
Crowd Flow
Auto-Encoder
Collaborative Filtering
Recommended Sites
(c)

Related Papers

- Wang, L., Geng, X., Ma, X., Liu, F., & Yang, Q. (2019). Cross-city transfer learning for deep spatio-temporal prediction. *IJCAI*.
- Wang, L., Geng, X., Ma, X., Zhang, D., & Yang, Q. (2019). Ridesharing Car Detection by Transfer Learning. *Artificial Intelligence*.
- Wang, L., Guo, B., & Yang, Q. (2018). Smart City Development with Urban Transfer Learning. *IEEE Computer*.
- Yao, H., Liu, Y., Wei, Y., Tang, X., & Li, Z. (2019). Learning from Multiple Cities: A Meta-Learning Approach for Spatial-Temporal Prediction. *In The World Wide Web Conference* (pp. 2181-2191).
- Guo, B., Li, J., Zheng, V. W., Wang, Z., & Yu, Z. (2018). CityTransfer: Transferring Inter-and Intra-City Knowledge for Chain Store Site Recommendation based on Multi-Source Urban Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (UbiComp), 1(4), 135*.
- Wei, Y., Zheng, Y., & Yang, Q. (2016). Transfer knowledge between cities. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1905-1914)*.
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science, 353(6301), 790-794*.
- Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2014). Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST), 5(3), 38*.
- Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering, 22(10), 1345-1359*.

Thanks !

leyewang@pku.edu.cn
https://sites.google.com/site/wangleye/