Worker Selection Towards Data Completion for Online Sparse Crowdsensing

Wenbin Liu*, En Wang*, Yongjian Yang*, and Jie Wu+

*Jinlin University, +Temple University

Outline

I. Background and Motivation
II. Problem and Framework
III. Data Completion
IV. Importance Estimation
V. Worker Selection
VI. Performance Evaluation
VII. Conclusion

Outline

I. **Background and Motivation**

II. Problem and Framework

III. Data Completion

IV. Importance Estimation

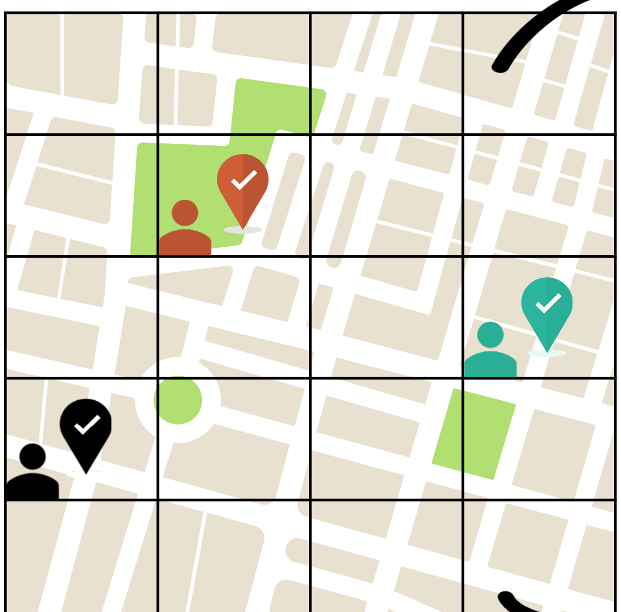
V. Worker Selection

VI. Performance Evaluation

VII. Conclusion

Mobile CrowdSensing (MCS)

- Recruit users with mobile devices to perform various sensing tasks



Traffic Monitoring

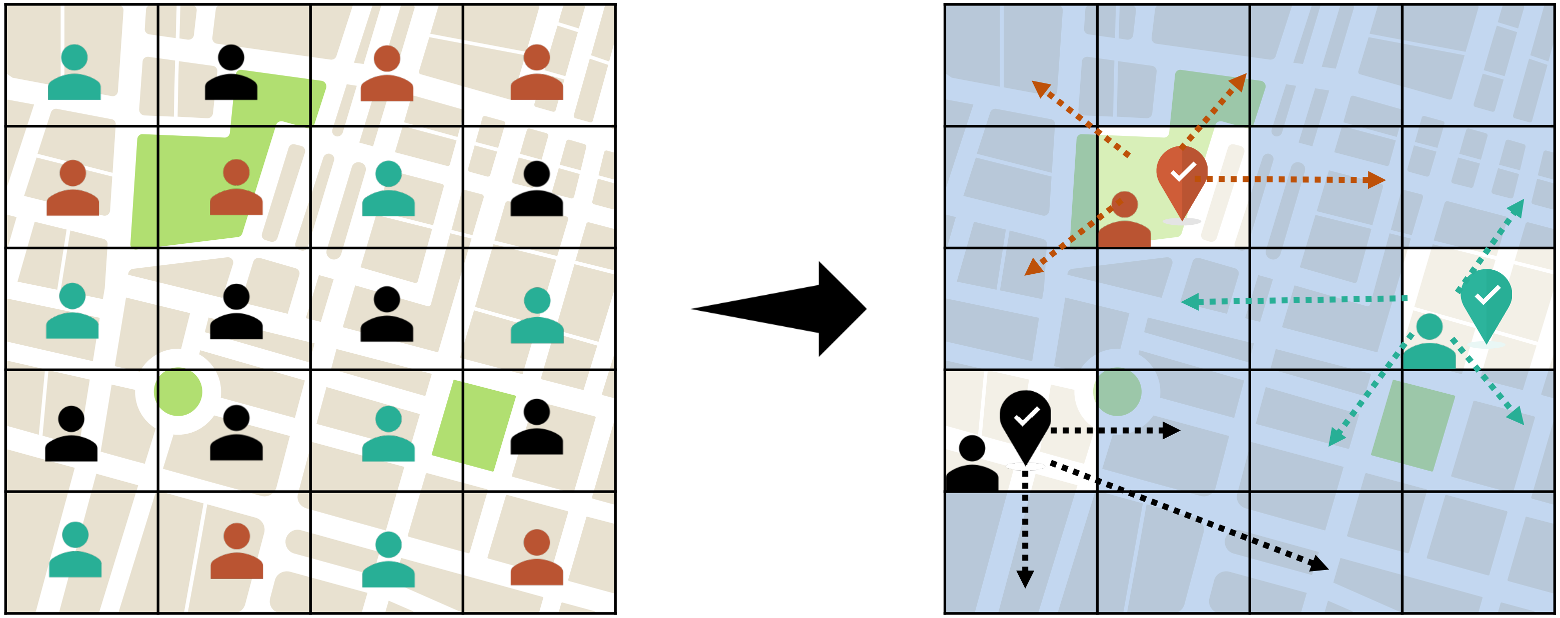
Air Quality Sensing

Crowdsourced Parking

... ...

Sparse MCS

- MCS: a large number of workers
- Sparse MCS: sense a few and infer the rest



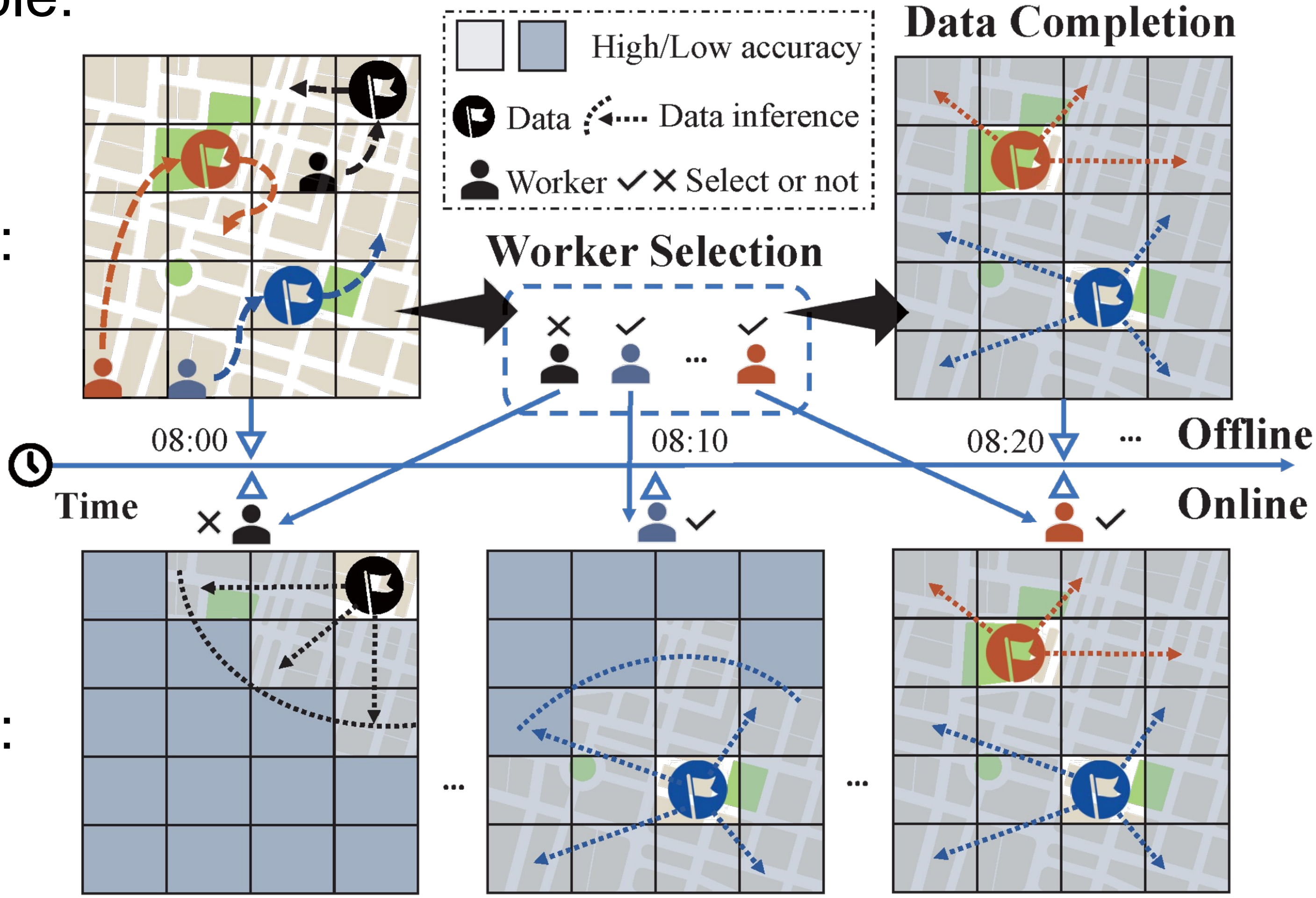
Offline vs. online

- Offline Sparse MCS
  - Pre-determined worker pool
  - Infer the rest after receiving all the data
- Online Sparse MCS
  - Workers participate in real time
  - Dynamically coming data
    - ✓ More realistic scenarios

Offline vs. online

- Example:

1. Offline:
2. Online:



Online Data Completion

- Intuitively: complete after receiving each new data
  - Cost a lot with high completion latency
- Alternatively: group data into batches
  - Still exist a lag between receiving and exploiting
- First challenge:
  - How to effectively exploit the dynamically coming data for **online data completion**?

Area Importance Estimation

- Some spatio-temporal areas are more important
  - Data from center areas ▶ the corner ones
- Area importance is time-varying
  - Newly obtained data ▶ the old ones
- Second challenge:
  - How to **estimate area importance** for improving data completion?

Online Worker Selection

- Instead of passively waiting for given data
  - Select workers to actively sense important areas for accurate data completion
- Online scenarios
  - Workers and data are invisible and hard to predict
- Third challenge:
  - How to select the **worker set** in an **online** manner to actively sense important area for accurate completion?

Outline

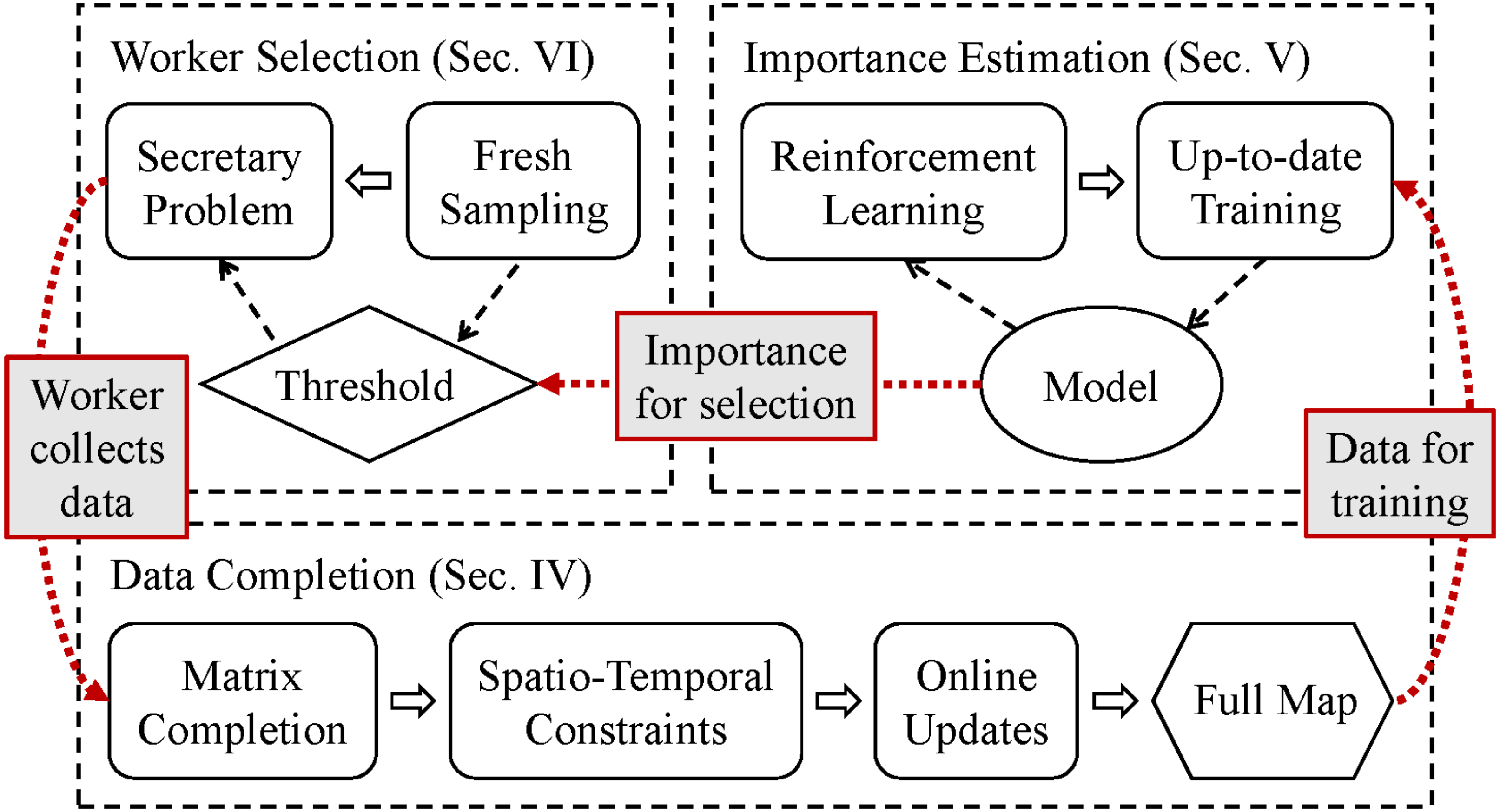
Problem Formulation

- **Problem** [Worker Selection Towards Data Completion for Online Sparse MCS]:
- Given a set of tasks with m sensing areas and n cycles, under a budget B and the duration T, our problem is to select a set of sequential participating workers μ
- Goal: minimizing the total completion error

$$\begin{aligned} \text{minimize} \quad & \sum_{t=1}^{T} \varepsilon(Y, \hat{Y}_t) \\ \text{subject to} \quad & \hat{Y}_t = f(Y_t'),\ \mu \subseteq W,\ \sum_{u_i \in \mu} c_i \leq B \end{aligned}$$

Framework Overview

- Online Sparse MCS (OS-MCS)



Outline

I. Background and Motivation

II. Problem and Framework

III. Data Completion

IV. Importance Estimation

V. Worker Selection

VI. Performance Evaluation

VII. Conclusion

Online Data Completion

- A. Matrix Completion

1. In the physical world, the sensing data naturally exist with some correlations (*low-rank matrix*)

$$\min \ \operatorname{rank}(\hat{Y}), \qquad \text{s.t.}, \hat{Y} \otimes M = Y'.$$

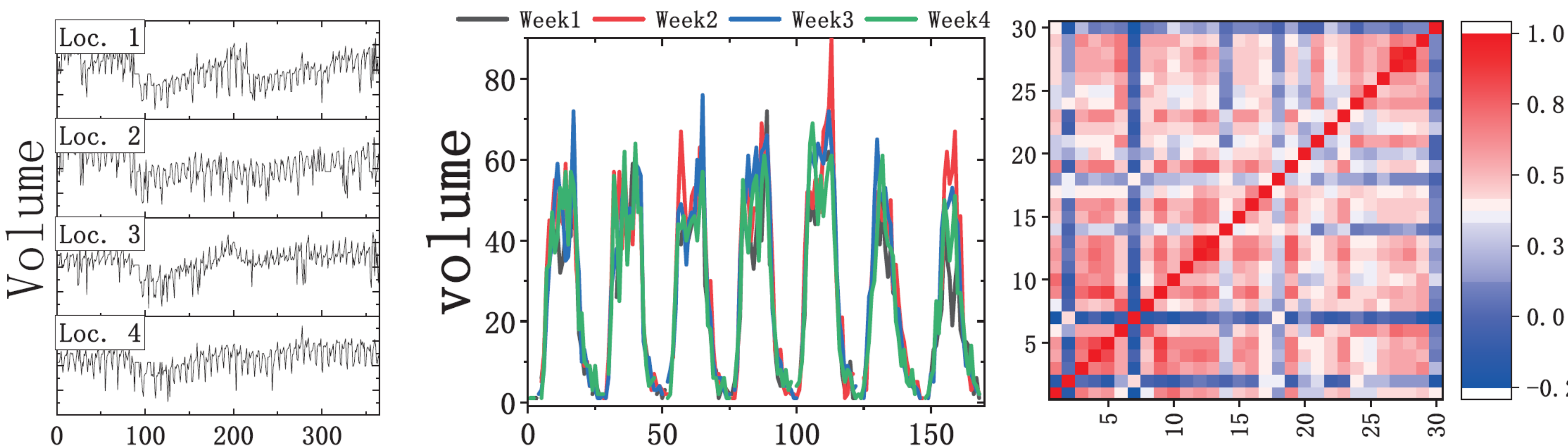
2. Factor the low-rank matrix into the latent spatio-temporal feature matrices

$$\hat{Y}_{m\times n} = U_{m\times r} V_{n\times r}^T$$

$$\min \ \|(Y' - UV^T) \otimes M\|_F^2 + \lambda(\|U\|_F^2 + \|V\|_F^2)$$

Online Data Completion

- B. Spatio-Temporal Constraints
  - ✓ Usually, there exist spatial and temporal correlations (*continuity, periodicity, and similarity*)

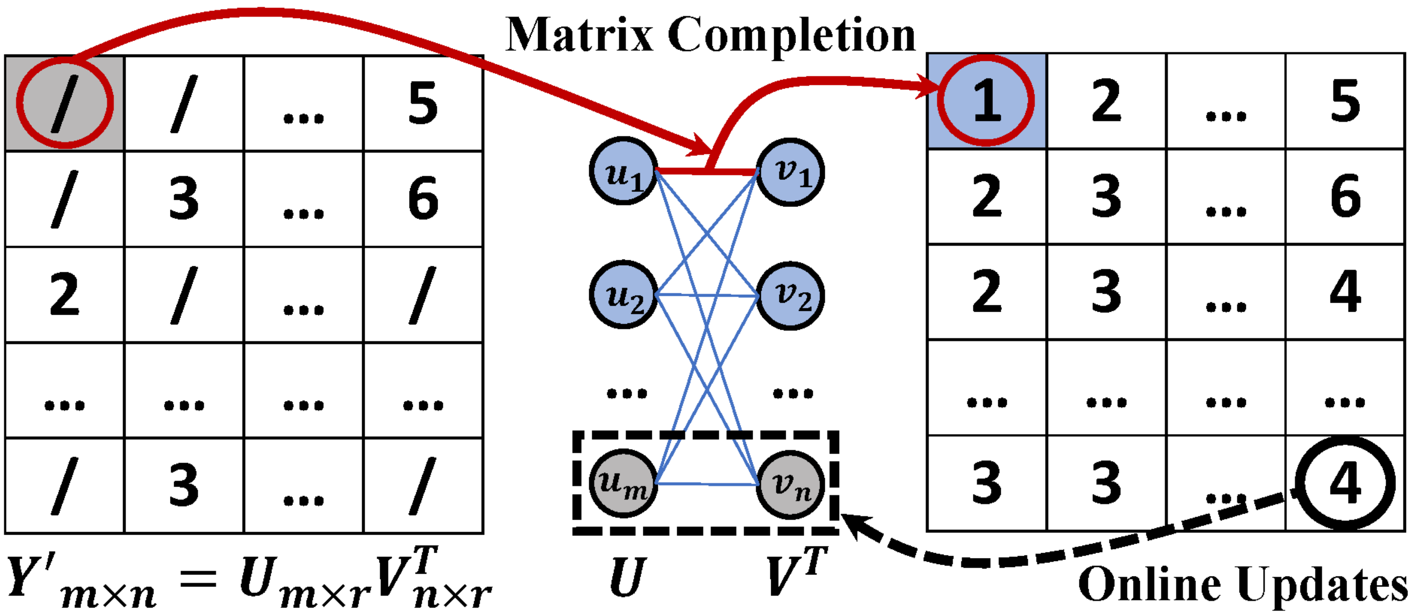


(a) Continuity (b) Periodicity (c) Similarity

$$\min \| (Y' - UV^T) \otimes M \|_F^2 + \lambda (\| U \|_F^2 + \| V \|_F^2) + g(\mathbb{T}) + h(\mathbb{S})$$

Online Data Completion

- C. Online Updates
  - ✓ With a new coming data, the new spatio-temporal matrices *U/V* are close to old ones (*update tall matrices*)



Outline

I. Background and Motivation

II. Problem and Framework

III. Data Completion

IV. Importance Estimation

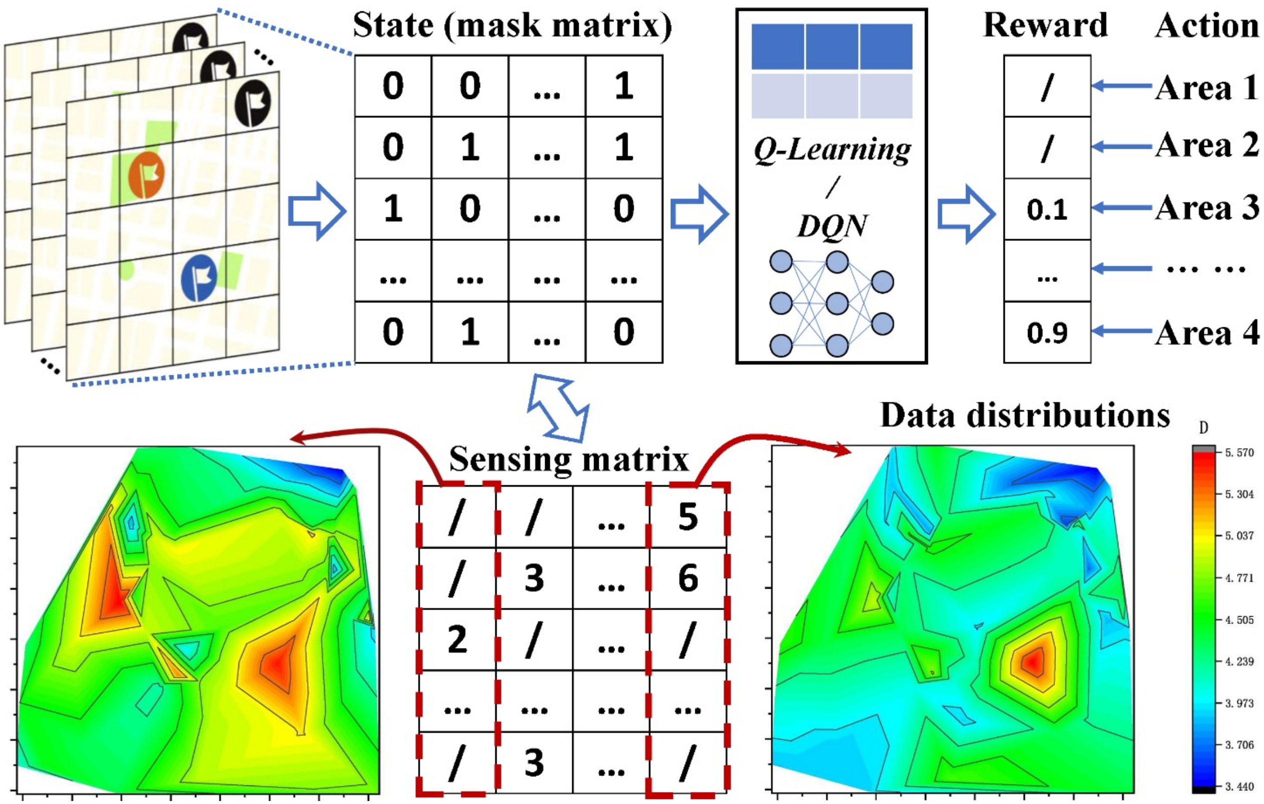
V. Worker Selection

VI. Performance Evaluation

VII. Conclusion

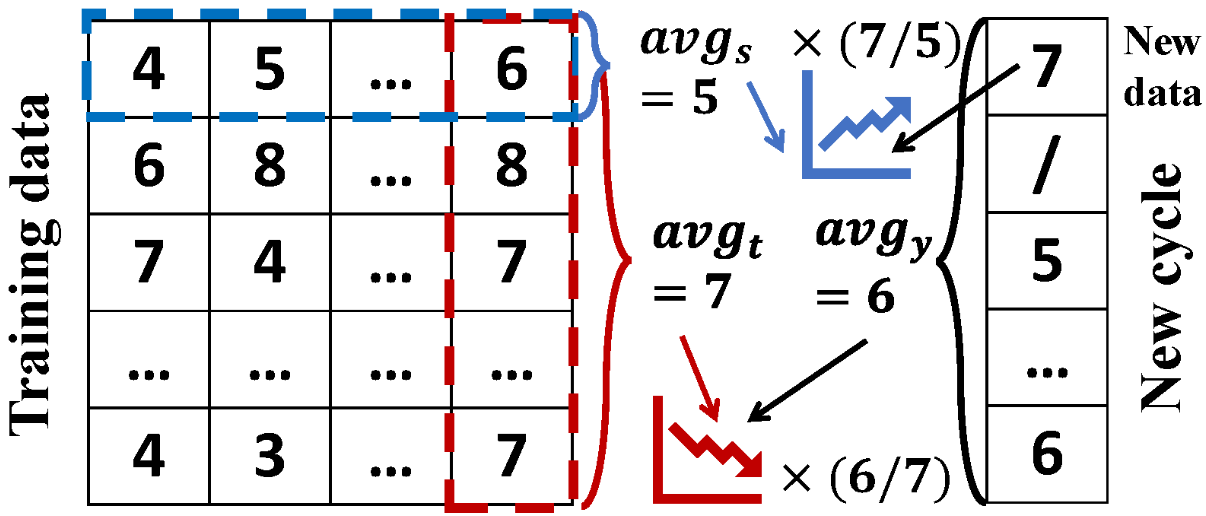
Area Importance Estimation

- A. Reinforcement Learning
  - ✓ Connect areas with completion accuracy directly



Area Importance Estimation

- B. Up-to-date Training
  - The sensing data may change a lot over time
    - ✓ Keep the model up-to-date
  - ➤ Utilize gradual changes in adjacent cycles and areas



Outline

I. Background and Motivation

II. Problem and Framework

III. Data Completion

IV. Importance Estimation

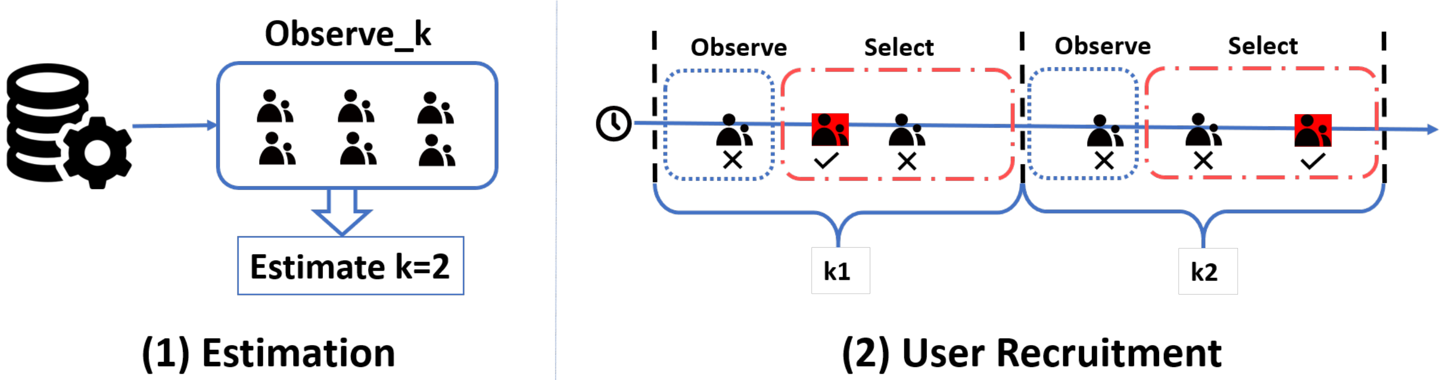
V. Worker Selection

VI. Performance Evaluation

VII. Conclusion

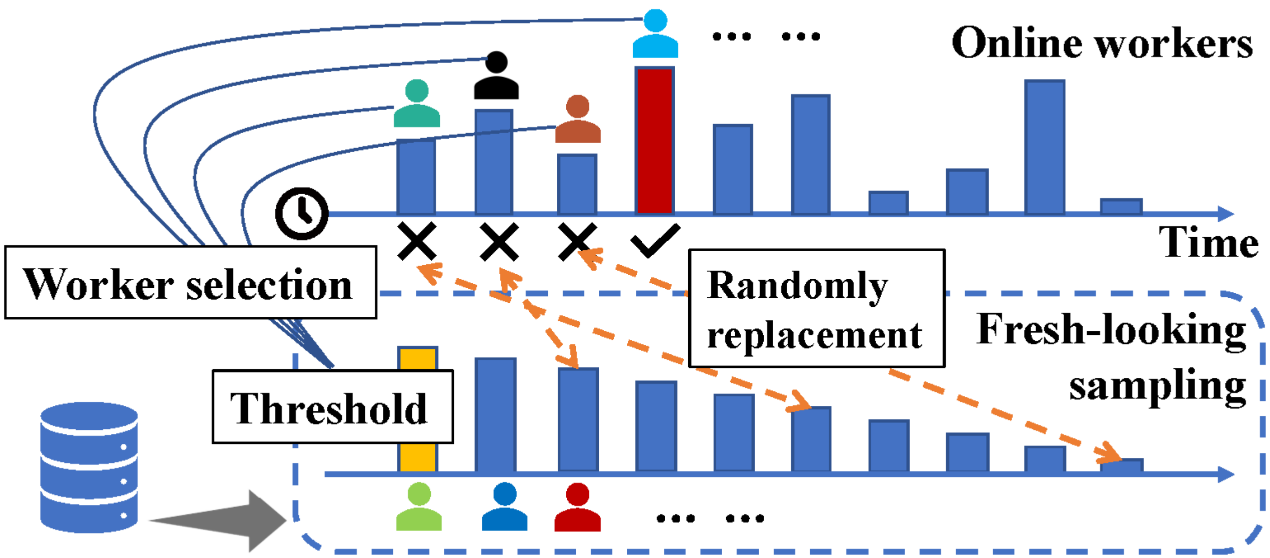
Online Worker Selection

- A. Secretary Problem-based Worker Selection
  1. Estimation k (*budget and time constraints*)
  2. Online Segmented Recruitment: non-submodular k-Secretaries Problem (*online recruiting process*)



Online Worker Selection

- B. Fresh-looking Sampling
  - Construct a sample worker set to approximate the first *1/e* workers[1]



[1] S. Ehsani, et al, "Prophet secretary for combinatorial auctions and matroids," in *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2018, pp. 700–714.

Outline

I. Background and Motivation

II. Problem and Framework

III. Data Completion

IV. Importance Estimation

V. Worker Selection

VI. Performance Evaluation

VII. Conclusion

Five typical sensing tasks

- Environmental monitoring
  - PM2.5[2], Temperature, Humidity[3]
- Urban sensing
  - Traffic[4], Parking[5]

	Environmental Monitoring			**Urban Sensing**	
	PM2.5	**Temperature**	**Humidity**	**Traffic**	**Parking**
City	Beijing (China)	Lausanne (Switzerland)		New South Wales (Australia)	Birmingham (UK)
Sensing areas	36 areas each with 1k*1km^2	57 areas each with 50*30m^2		30 subway stations	73 car parks
Cycle & Duration	1 hour & 11 days	0.5 hour & 7 days		1 day & 1 year	0.5 hour & 77 days
Mean Std.	79.11 ± 81.21	$6.04 \pm 1.87°C$	$84.52 \pm 6.32\%$	19095.73 ± 26750.79	647.97 ± 657.23

[2] Y. Zheng, et al, "U-air:when urban air quality inference meets big data," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2013, pp. 1436–1444.

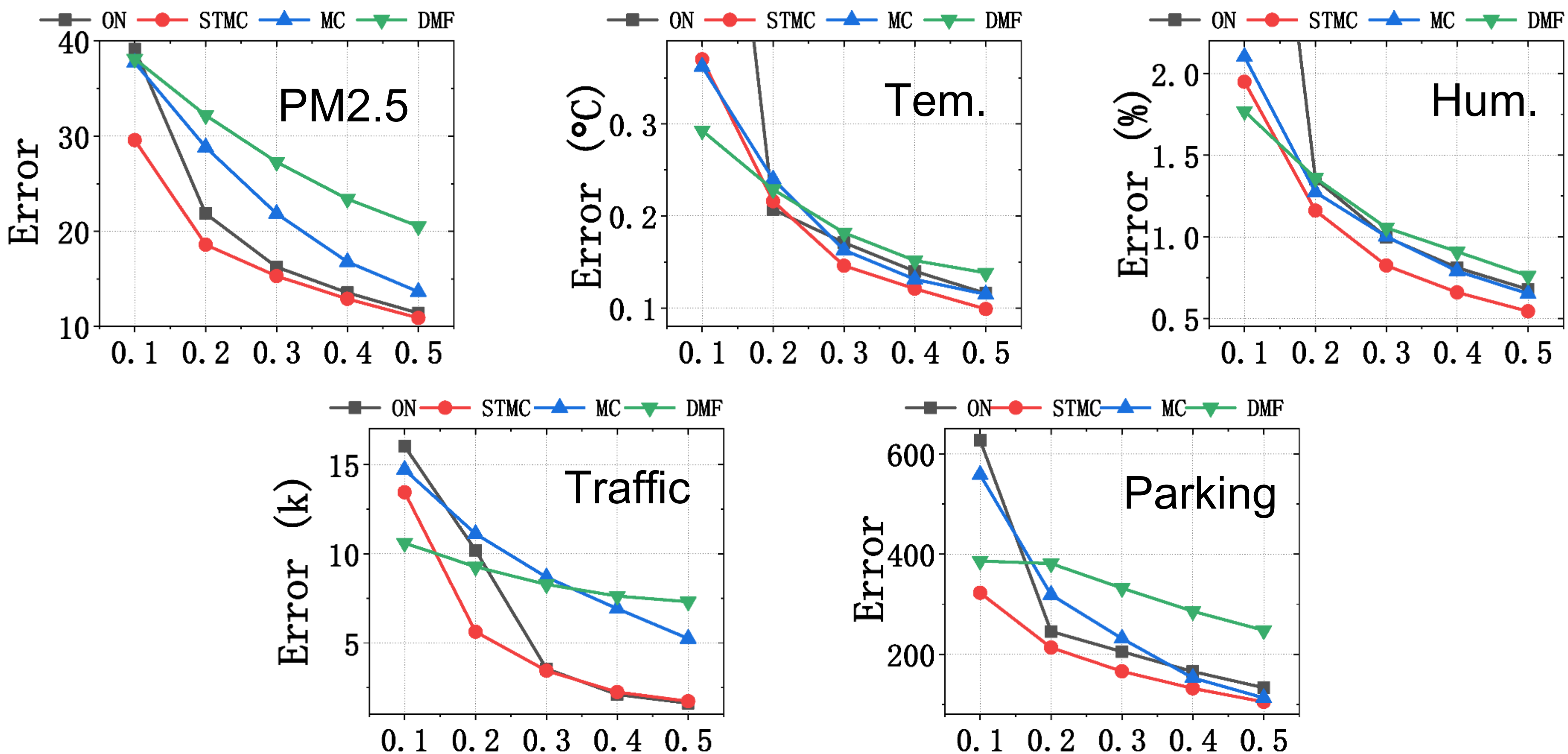
[3] F. Ingelrest, et al, "Sensorscope: Application-specific sensor network for environmental monitoring," *ACM Transactions on Sensor Networks*, vol. 6, no. 2, pp. 1–32, 2010.

[4] T. for NSW. (2019) System of traffic volume viewer. @ONLINE. [Online]. Available: https://www.rms.nsw.gov.au/about/corporatepublications/statistics/traffic-volumes/aadt-map/index.html

[5] D. H. Stolfi, et al, "Predicting car park occupancy rates in smart cities," in *International Conference on Smart Cities*. Springer, 2017, pp. 107–117.

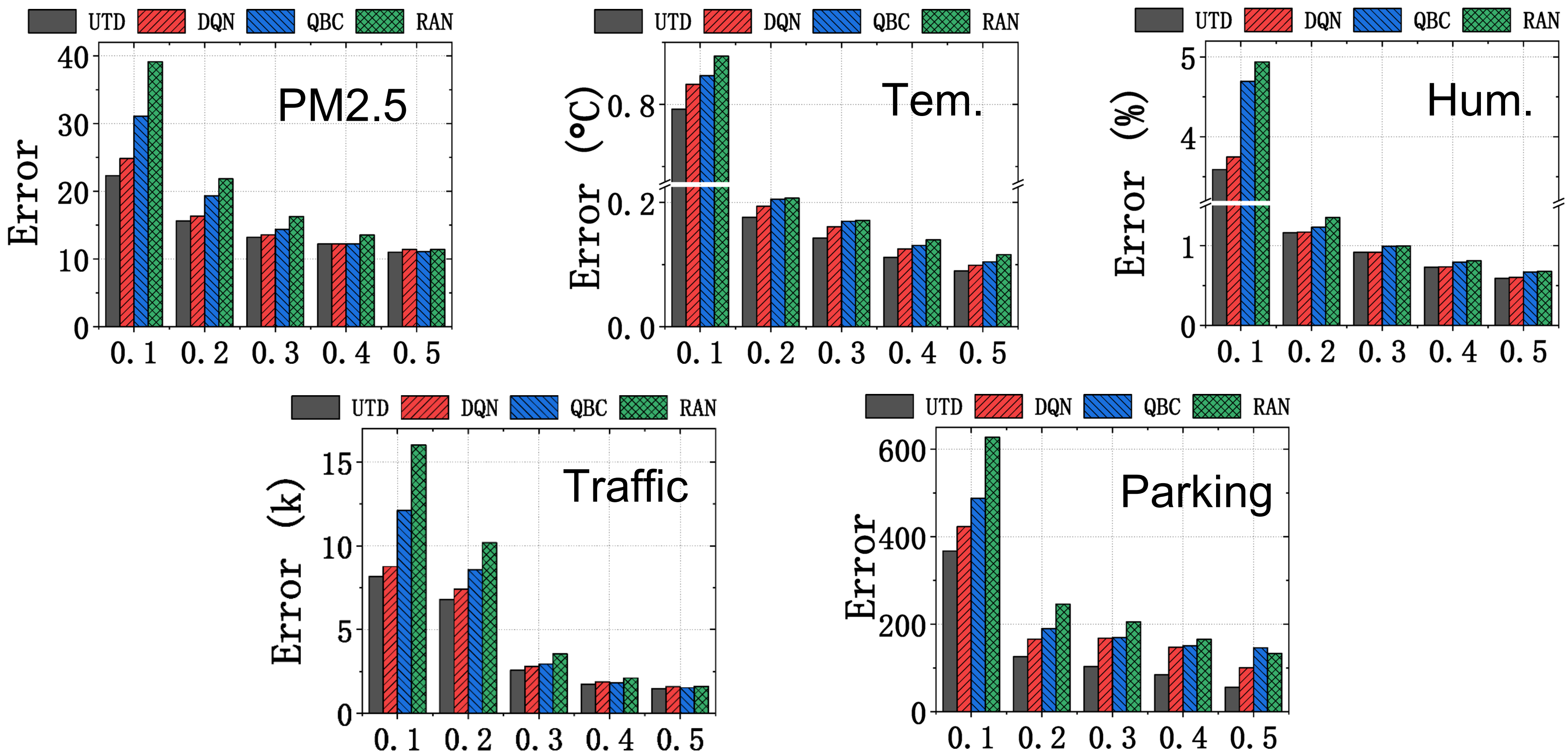
Main results

- Data completion under randomly selection



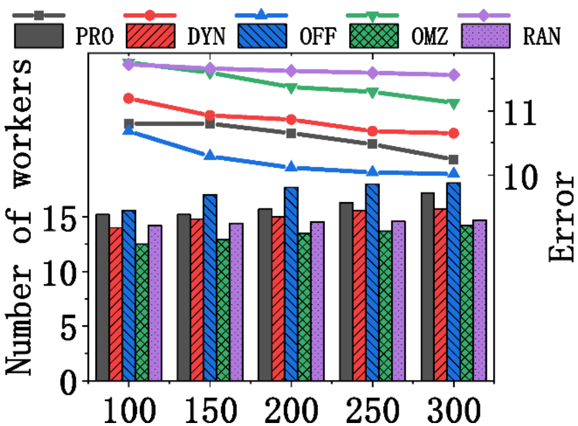
Main results

- Importance estimation-guided data completion

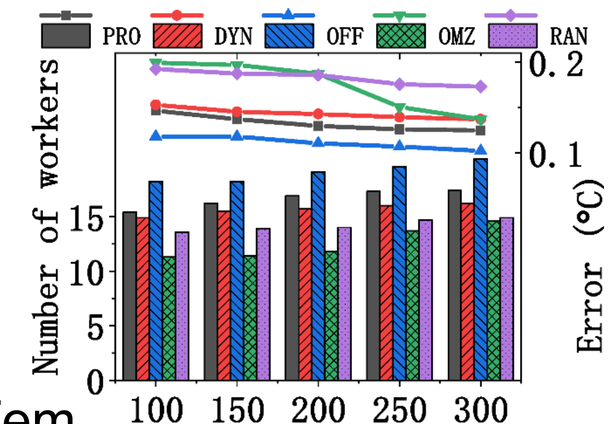


Main results

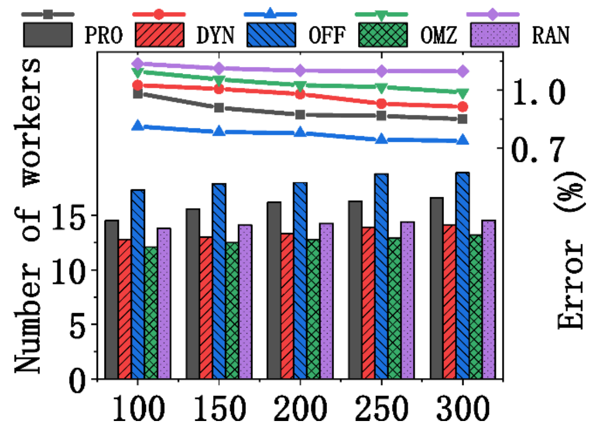
- Worker selection towards data completion



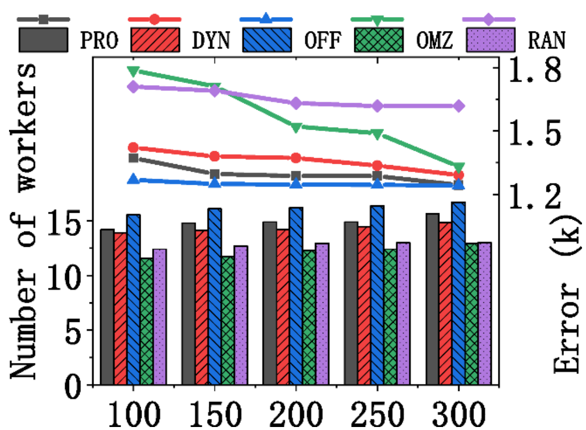
PM2.5



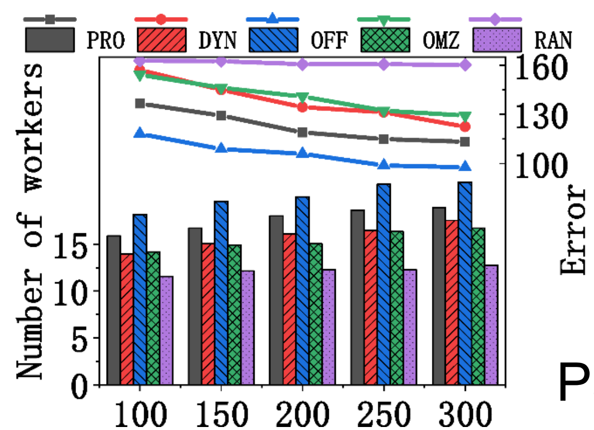
Tem.



Hum.



Traffic



Parking

Main results

- Spatio-temporal weights and running time:

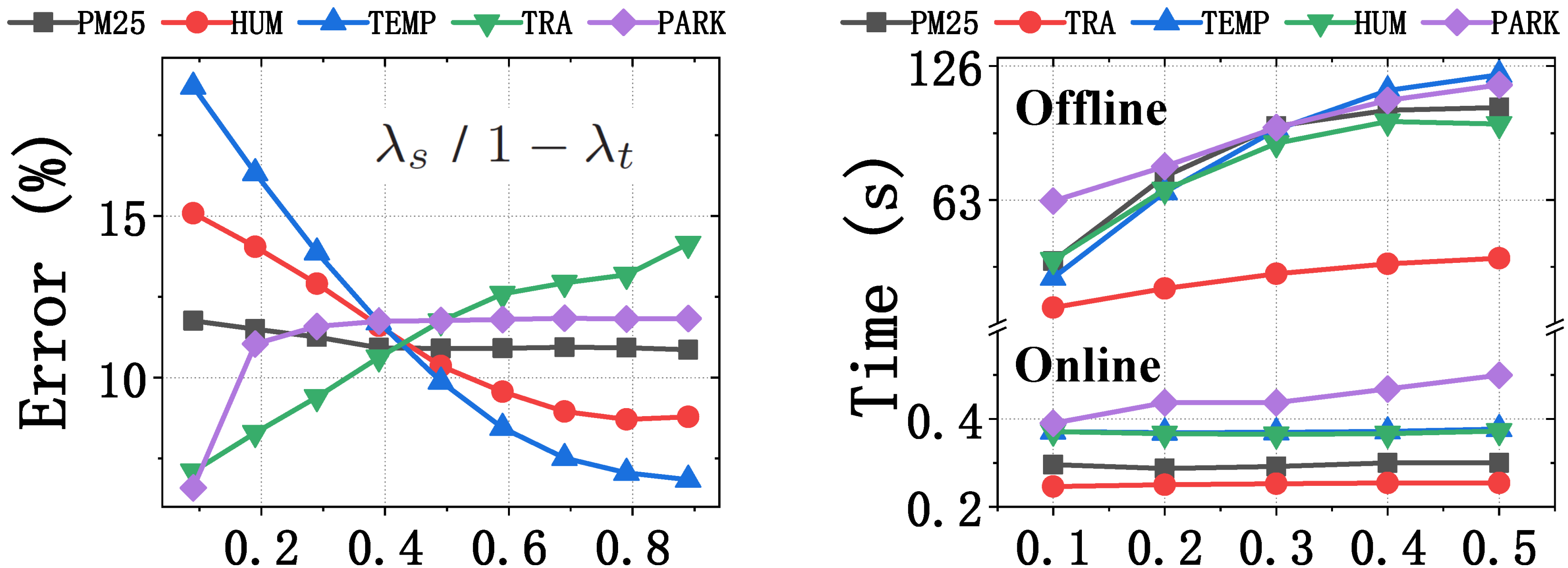


TABLE III: Running time

ON	STMC	MC	DMF	KNN-S	KNN-T	GP
0.32	73.87	55.19	24.81	0.05	0.05	0.03

Outline

I. Background and Motivation

II. Problem and Framework

III. Data Completion

IV. Importance Estimation

V. Worker Selection

VI. Performance Evaluation

VII. Conclusion

Conclusion

- *Online Sparse Crowdsensing*:
  - Select suitable workers to actively sense important areas for online data completion
- *Framework OS-MCS*:
  - Matrix Completion with Online Updates
  - RL-based Estimation with Up-to-date Training
  - Online Selection with Fresh-looking Sampling
- *Extensive Evaluation*
  - Five typical sensing tasks (in two types)

Q&A

Thank you!

Q&A