Exploiting Outlier Value Effects in Sparse Urban CrowdSensing

吉林大学计算机科学与技术学院
College of Computer Science and Technology, Jilin University

College of Science and Technology
TEMPLE UNIVERSITY

En Wang, Mijia Zhang,
Yongjian Yang, Yuanbo Xu*

Jie Wu

Speaker: Mijia Zhang

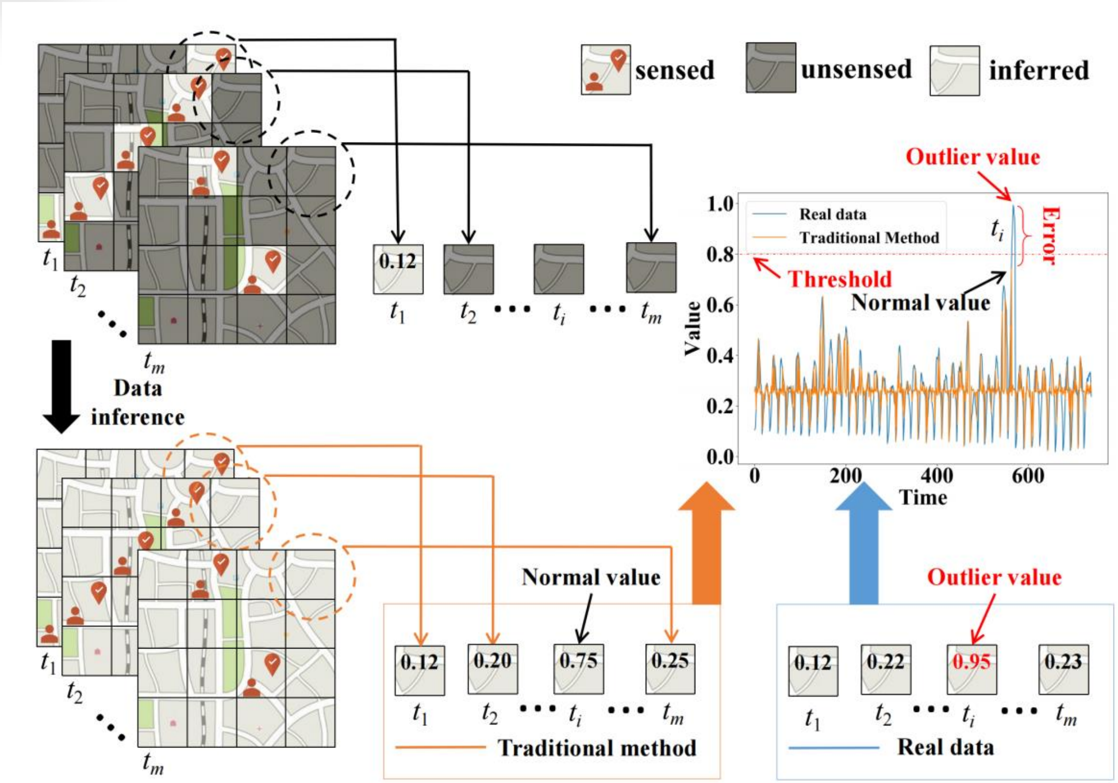
IWQoS 2021

Outlines

1. Introduction and challenges
2. Contributions
3. Model
4. Experiments

Section 1: Introduction and challenges

1. Introduction and challenges



An example to describe the outlier value effect in data inference problem.

1) rarity and unpredictability

2) inconsistency compared to normal values

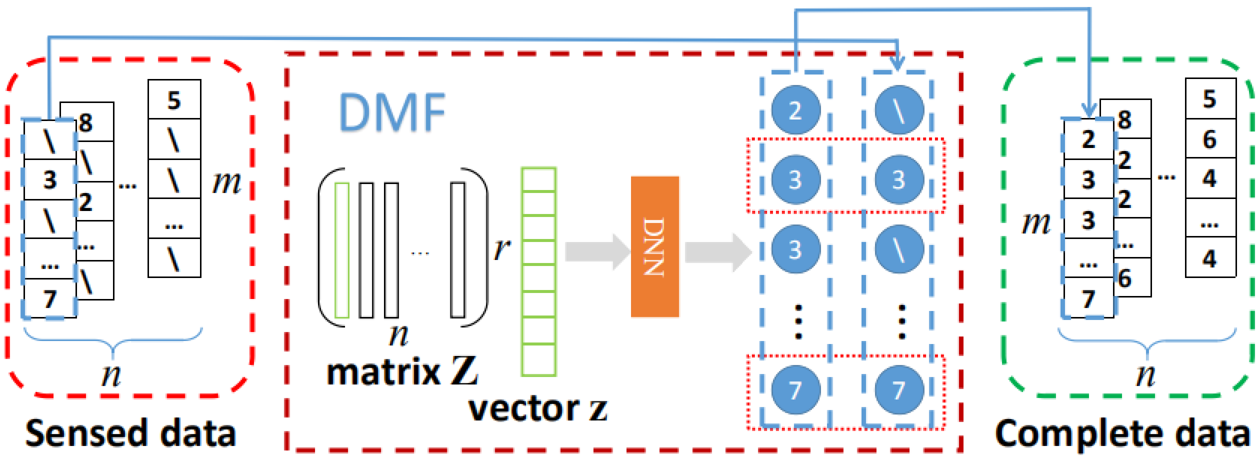
3) complex spatiotemporal relations

1. Introduction and challenges

Challenges

- recover outlier values from such rare outlier value data
- deal with inconsistent data distribution
- extract the complex spatiotemporal relationship of outlier value data

- How to deal with challenges
  - DMF + OVLoss



  - Outlier Value Model

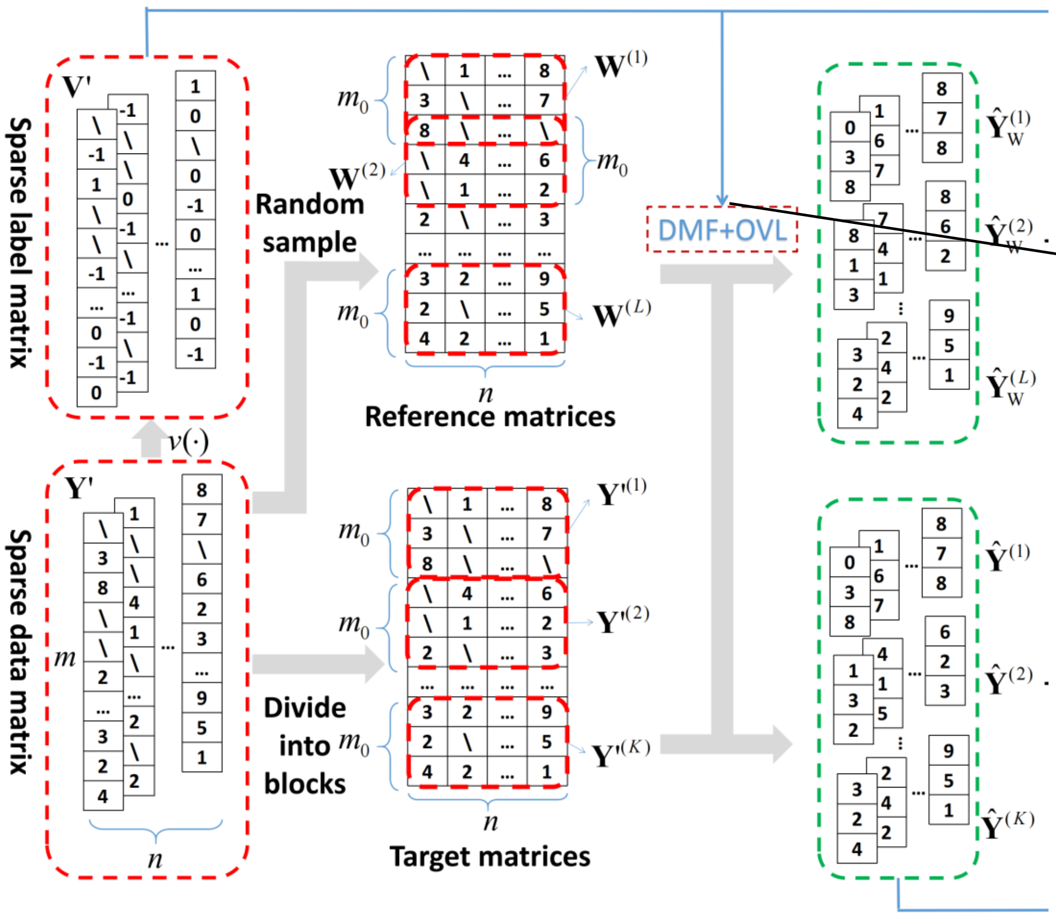
Section 2: Contributions

2. Contributions

- Our work has the following contributions:
  - formalize the sparse urban crowdsensing problem
  - propose an urban crowdsensing method named DMF-OV
  - evaluate the proposed method on three real-world datasets with three typical urban sensing tasks.

Section 3: Model

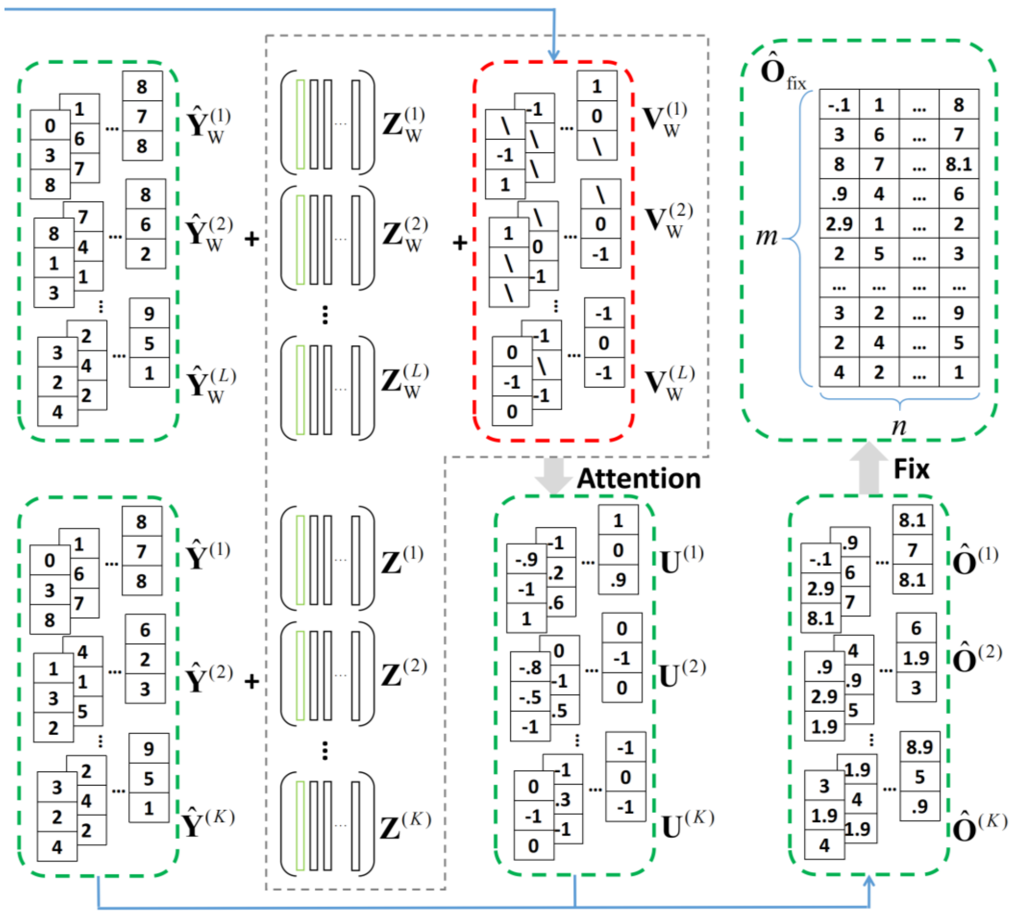
Framework Overview (DMF + OVLoss)



$$L_{\text{mix}} = \xi L_{\text{MSE}} + (1-\xi) L_{\text{OVL}}$$

3. Model

❖ Framework Overview (Outlier Value Model)



- Embedding Module

$$\mathbf{Z}_{\mathrm{W}} = \{\mathbf{Z}_{\mathrm{W}}^{(1)}, \mathbf{Z}_{\mathrm{W}}^{(2)}, \cdots, \mathbf{Z}_{\mathrm{W}}^{(L)}\}$$

- Label Matrix Module

$$\mathbf{V}_{\mathrm{W}} = \{\mathbf{V}_{\mathrm{W}}^{(1)}, \mathbf{V}_{\mathrm{W}}^{(2)}, \cdots, \mathbf{V}_{\mathrm{W}}^{(L)}\}$$

Section 4: Experiments

4. Experiments

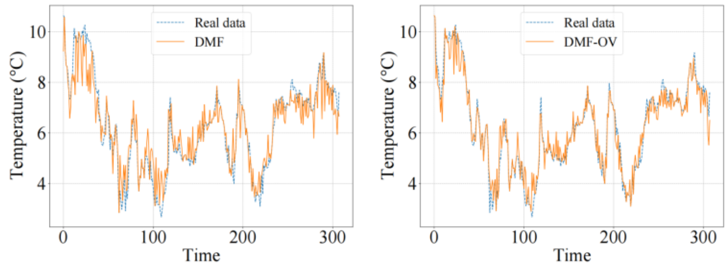
❖ Statistics of three evaluation datasets.

	Datasets		
	Sensor-Scope	*U-Air*	*Parking in Birmingham*
City	Lausanne (Switzerland)	Beijing (China)	Birmingham (UK)
Data	Temperature	PM2.5	Parking occupancy rate
Subarea	57 subareas each with $50 \times 30m^2$	36 subareas each with $1000 \times 1000m^2$	30 parking lots
Period & Duration	$0.5h$ & $7d$	$1h$ & $11d$	$0.5h$ & $77d$
Mean $\pm$ Std. (Unit)	6.04 ± 1.87 (°C)	79.11 ± 81.21 (μg/m^3)	53.6 ± 26.3 (%)

- KNN
- GP
- DMF
- IGMC

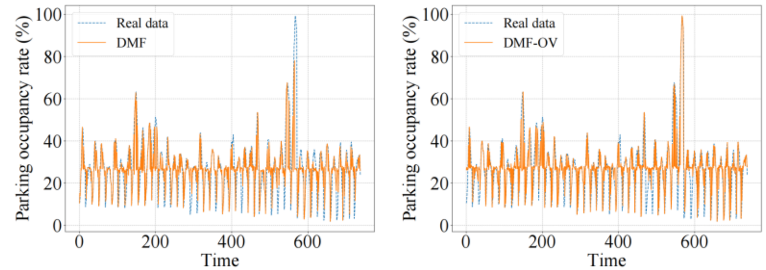
4. Experiments

- RQ1: Does our method really work for outlier value data effectively?



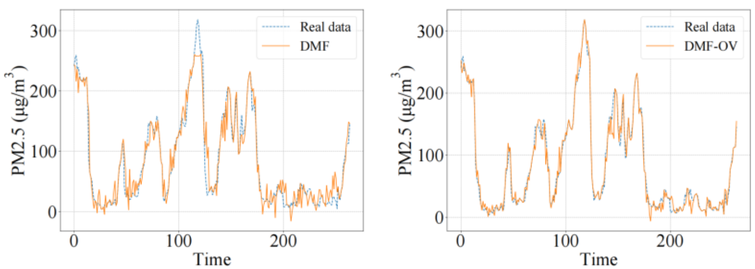
(a) DMF (b) DMF-OV

Fig. 4: Complementary effects of outlier values over *Sensor-Scope*.



(a) DMF (b) DMF-OV

Fig. 6: Complementary effects of outlier values over *Parking in Birmingham*.



(a) DMF (b) DMF-OV

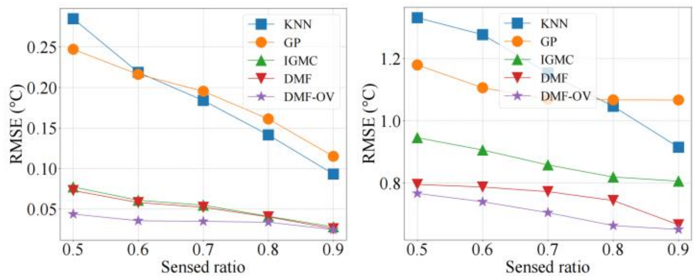
Fig. 5: Complementary effects of outlier values over *U-Air*.

TABLE III: RMSE of outlier values over all three tasks.

	DMF	DMF-OV
Temperature ($^\circ$C)	0.67	0.45
PM2.5 ($\mu g/m^3$)	20.56	11.74
Parking occupancy rate (%)	10.4	8.9

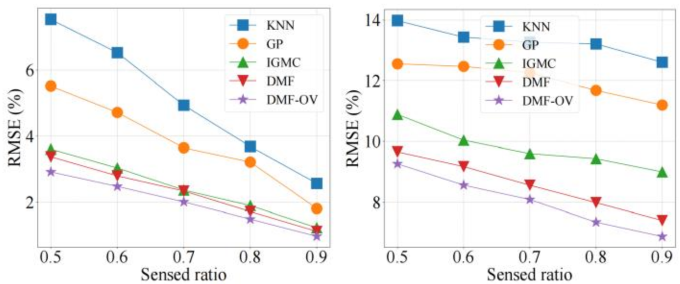
4. Experiments

- RQ2: Does our method improve the accuracy of matrix completion and prediction?



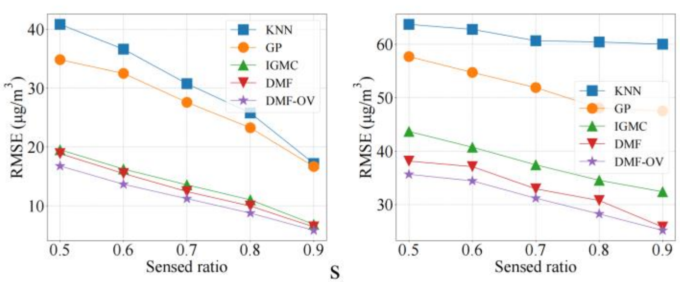
(a) Inference task (b) Prediction task

Fig. 7: Inference and prediction accuracy under different sensed ratios over *Sensor-Scope*.



(a) Inference task (b) Prediction task

Fig. 9: Inference and prediction accuracy under different sensed ratios over *Parking in Birmingham*.

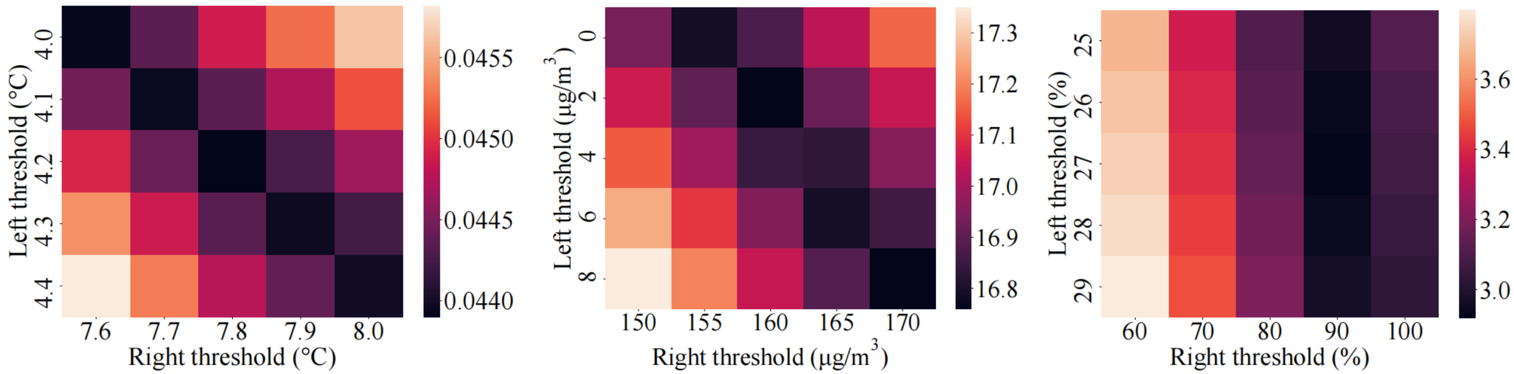


(a) Inference task (b) Prediction task

Fig. 8: Inference and prediction accuracy under different sensed ratios over *U-Air*.

4. Experiments

- RQ3: What are the influences of hyper-parameters in the model?



(a) Temperature (b) PM2.5 (c) Occupancy Rate

Fig. 10: Inference accuracy under different hyper-parameters over *Sensor-Scope*, *U-Air*, and *Parking in Birmingham*.

Thanks for listening.

Q&A