Exploiting Outlier Value Effects in Sparse Urban CrowdSensing

Speaker: Mijia Zhang

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
1. Introduction and challenges

- 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
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**
  \[
  Z_W = \{ Z_W^{(1)}, Z_W^{(2)}, \cdots, Z_W^{(L)} \}
  \]

- **Label Matrix Module**
  \[
  V_W = \{ V_W^{(1)}, V_W^{(2)}, \cdots, V_W^{(L)} \}
  \]
Section 4: Experiments
4. Experiments

Statistics of three evaluation datasets.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Sensor-Scope</th>
<th>U-Air</th>
<th>Parking in Birmingham</th>
</tr>
</thead>
<tbody>
<tr>
<td>City</td>
<td>Lausanne (Switzerland)</td>
<td>Beijing (China)</td>
<td>Birmingham (UK)</td>
</tr>
<tr>
<td>Data</td>
<td>Temperature</td>
<td>PM2.5</td>
<td>Parking occupancy rate</td>
</tr>
<tr>
<td>Subarea</td>
<td>57 subareas each with $50 \times 30m^2$</td>
<td>36 subareas each with $1000 \times 1000m^2$</td>
<td>30 parking lots</td>
</tr>
<tr>
<td>Period &amp; Duration</td>
<td>0.5h &amp; 7d</td>
<td>1h &amp; 11d</td>
<td>0.5h &amp; 77d</td>
</tr>
<tr>
<td>Mean ± Std. (Unit)</td>
<td>6.04 ± 1.87 ($^\circ$C)</td>
<td>79.11 ± 81.21 ($\mu g/m^3$)</td>
<td>53.6 ± 26.3 (%)</td>
</tr>
</tbody>
</table>

- KNN
- GP
- DMF
- IGMC
4. Experiments

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

Fig. 4: Complementary effects of outlier values over *SensorScope.*

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

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

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

<table>
<thead>
<tr>
<th></th>
<th>DMF</th>
<th>DMF-OV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature (°C)</td>
<td>0.67</td>
<td>0.45</td>
</tr>
<tr>
<td>PM2.5 (µg/m³)</td>
<td>20.56</td>
<td>11.74</td>
</tr>
<tr>
<td>Parking occupancy rate (%)</td>
<td>10.4</td>
<td>8.9</td>
</tr>
</tbody>
</table>
4. Experiments

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

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

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

Fig. 9: Inference and prediction accuracy under different sensed ratios over Parking in Birmingham.
RQ3: What are the influences of hyper-parameters in the model?

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

Q&A