Worker Selection Towards Data Completion for Online Sparse Crowdsensing

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

*Jinlin University, +Temple University
I. Background and Motivation
II. Problem and Framework
III. Data Completion
IV. Importance Estimation
V. Worker Selection
VI. Performance Evaluation
VII. Conclusion
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

May 2022
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
Example:

1. Offline:

2. Online:
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?
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?
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?
I. Background and Motivation
II. Problem and Framework
III. Data Completion
IV. Importance Estimation
V. Worker Selection
VI. Performance Evaluation
VII. Conclusion
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 $\mu$

Goal: minimizing the total completion error

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

Online Sparse MCS (OS-MCS)

Worker Selection (Sec. VI)
- Secretary Problem
- Fresh Sampling

Importance Estimation (Sec. V)
- Reinforcement Learning
- Up-to-date Training

Data Completion (Sec. IV)
- Matrix Completion
- Spatio-Temporal Constraints
- Online Updates

Worker collects data
Threshold
Importance for selection
Model
Data for training
Full Map

May 2022
I. Background and Motivation
II. Problem and Framework
III. Data Completion
IV. Importance Estimation
V. Worker Selection
VI. Performance Evaluation
VII. Conclusion
A. Matrix Completion

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

   \[ \min \ \text{rank}(\hat{Y}), \ \text{s.t.}, \ \hat{Y} \otimes M = Y'. \]

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

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

   \[ \min \ \|(Y' - UV^T) \otimes M\|^2_F + \lambda(\|U\|^2_F + \|V\|^2_F) \]
B. Spatio-Temporal Constraints

- Usually, there exist spatial and temporal correlations (continuity, periodicity, and similarity)

\[
\min \| (Y' - U V^T) \otimes M \|_F^2 + \lambda (\| U \|_F^2 + \| V \|_F^2) + g(T) + h(S)
\]
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
A. Reinforcement Learning

Connect areas with completion accuracy directly

State (mask matrix)

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>...</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>...</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>...</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Reward

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/</td>
<td>/</td>
<td>/</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Action

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Area 1</td>
<td>Area 2</td>
<td>Area 3</td>
<td>Area 4</td>
</tr>
</tbody>
</table>

Data distributions

Sensing matrix

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/</td>
<td>/</td>
<td>...</td>
<td>5</td>
</tr>
<tr>
<td>/</td>
<td>3</td>
<td>...</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>/</td>
<td>...</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>/</td>
<td>3</td>
<td>...</td>
<td>/</td>
</tr>
</tbody>
</table>
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
I. Background and Motivation
II. Problem and Framework
III. Data Completion
IV. Importance Estimation
V. Worker Selection
VI. Performance Evaluation
VII. Conclusion
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*)
B. Fresh-looking Sampling

- Construct a sample worker set to approximate the first $1/e$ workers\[^1\]

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\(^5\)

<table>
<thead>
<tr>
<th>City</th>
<th>PM2.5 Sensing areas</th>
<th>Temperature</th>
<th>Humidity</th>
<th>City</th>
<th>Traffic</th>
<th>Parking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beijing (China)</td>
<td>36 areas each with 1k*1km(^2)</td>
<td>0.5 hour &amp; 7 days</td>
<td>84.52 ± 6.32%</td>
<td>New South Wales (Australia)</td>
<td>30 subway stations</td>
<td>73 car parks</td>
</tr>
<tr>
<td>Lausanne (Switzerland)</td>
<td>57 areas each with 50*30m(^2)</td>
<td>1 day &amp; 1 year</td>
<td>19095.73 ± 26750.79</td>
<td>Birmingham (UK)</td>
<td>0.5 hour &amp; 77 days</td>
<td>647.97 ± 657.23</td>
</tr>
<tr>
<td>Cycle &amp; Duration</td>
<td>Mean Std.</td>
<td>6.04 ± 1.87°C</td>
<td>79.11 ± 81.21</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Main results

- Data completion under randomly selection

**PM2.5**

**Tem.**

**Hum.**

**Traffic**

**Parking**

May 2022
Main results

Importance estimation-guided data completion

- PM2.5
- Temperature (°C)
- Humidity (%)
- Traffic
- Parking

Error:

- UTD
- DQN
- QBC
- RAN

May 2022
Main results

- Worker selection towards data completion
Main results

- Spatio-temporal weights and running time:

![Graphs showing spatio-temporal weights and running time](image)

<table>
<thead>
<tr>
<th>TABLE III: Running time</th>
</tr>
</thead>
<tbody>
<tr>
<td>ON</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>0.32</td>
</tr>
</tbody>
</table>
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)
Thank you!