Privacy-Preserving Online Task Assignment in Spatial Crowdsourcing: A Graph-based Approach

Hengzhi Wang, En Wang*, Yongjian Yang, Jie Wu, Falko Dressler

Speaker: Hengzhi Wang
Outlines

1. Introduction
2. Challenges and contributions
3. Problem formulation
4. Approaches
5. Experiments
1. Introduction

2. Challenges and contributions

3. Problem formulation

4. Approaches

5. Experiments
1. Introduction

Spatial Crowdsourcing

1) Background

A new problem-solving paradigm to explore the power of crowd with location-aware tasks.

2) Applications

Uber  Gigwalk  TaskRabbit
1. Introduction

Spatial Crowdsourcing

3) Components

- Tasks
- Platform
- Workers

4) Issues

- Task assignment
- Incentive mechanism
- Privacy protection
### Task assignment:

<table>
<thead>
<tr>
<th>Prior works</th>
<th>Our work</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) No-privacy</td>
<td>1) Privacy</td>
</tr>
<tr>
<td>2) One-worker-one-task</td>
<td>2) One-worker-many-tasks</td>
</tr>
<tr>
<td>3) Offline</td>
<td>3) Online</td>
</tr>
</tbody>
</table>

1. Location obfuscation
   - Worker and Task side
   - Real locations
   - Obfuscated locations
2. Task assignment
   - Platform side
   - One-worker-one-task
   - One-worker-many-tasks
3. Time
   - Initial time
   - End time
Outlines

1. Introduction
2. Challenges and contributions
3. Problem formulation
4. Approaches
5. Experiments
2. Challenges and contributions

- **Challenges**
  - Balance the tradeoff between privacy protection and utility
  - Execute the *one-worker-many-tasks* assignment
  - Deal with the *online* task assignment
2. Challenges and contributions

- Existing works cannot deal with these challenges

<table>
<thead>
<tr>
<th></th>
<th>1) Privacy</th>
<th>2) One-worker-many-tasks</th>
<th>3) Online</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work [1]</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Works [2,3]</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>Works [4,5]</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Our work</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

2. Challenges and contributions

- How to deal with challenges
  - tradeoff between privacy protection and utility
  - One-worker-many-tasks assignment
  - Online task assignment

Geo-Indistinguishability

PLP

Bound of $O(1/(\bar{d} \cdot \epsilon^2))$

TOA

Bound of $\frac{\Phi \cdot \delta}{d(0') - \sigma}$
2. Challenges and contributions

Our work has the following contributions:

- Propose a privacy mechanism to balanced the tradeoff between privacy and utility.

- Solve the online one-worker-many-tasks assignment problem with the competitive ratio of $O(1/\bar{d} \cdot \epsilon^2)$.

- Evaluate the effectiveness of the proposed method using real-world datasets.
1. Introduction
2. Challenges and contributions
3. Problem formulation
4. Approaches
5. Experiments
3. Problem formulation

- Privacy model

Geo-Indistinguishability [1]: A privacy mechanism $M$ satisfies Geo-Indistinguishability iff

$$M(l)(\mathcal{Z}) \leq e^{\epsilon d(l,l')} M(l')(\mathcal{Z})$$

- Privacy-preserving Online Task Assignment (POTA) Problem

Minimize $\sum_{w_i \in \mathcal{W}} d_i$

Subject to $\sum_{w_i \in \mathcal{W}} |T_i| \geq \delta$

$|T_i| \leq w_i \cdot C, \forall w_i \in \mathcal{W}$.

Outlines

1. Introduction
2. Challenges and contributions
3. Problem formulation
4. Approaches
5. Experiments
4. Approaches

- Planar Laplace distribution based Privacy mechanism (PLP)

**Algorithm 1: PLP**

<table>
<thead>
<tr>
<th>Input:</th>
<th>Privacy budget $\epsilon$, real location $l = (l_x, l_y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output:</td>
<td>Obfuscated location $l^* = (l_x^<em>, l_y^</em>)$</td>
</tr>
</tbody>
</table>

1. Draw $p \in [0, 1]$ uniformly, $\rho = F^{-1}(p)$;
2. Draw $\theta \in [0, 2\pi]$ uniformly;
3. $l_x^* = l_x + \rho \cos(\theta)$, $l_y^* = l_y + \rho \sin(\theta)$;
4. return $l^* = (l_x^*, l_y^*)$

**Planar Laplace distribution**

$$f(l, l^*, \epsilon) = \frac{\epsilon^2}{2\pi} e^{-\epsilon \cdot d(l, l^*)} \quad F(\rho, \epsilon) = \int_0^\rho \int_0^{2\pi} \frac{\epsilon^2}{2\pi} e^{-\epsilon \rho} d\theta d\rho = 1 - (1 + \epsilon \rho)e^{-\epsilon \rho}$$

$$l^* = (l_x + \Delta x, l_y + \Delta y) = (l_x + \rho \cos(\theta), l_y + \rho \sin(\theta))$$

**Theorem 1. Geo-Indistinguishability**

$$\mathcal{M}(l_1)(l^*)/\mathcal{M}(l_2)(l^*) = f(l_1, l^*, \epsilon)/f(l_2, l^*, \epsilon)$$

$$= e^{\epsilon \cdot (d(l_2,l^*) - d(l_1,l^*))} \leq e^{\epsilon \cdot d(l_1,d_2)}.$$
4. Approaches

- **Threshold-based Online task Assignment mechanism (TOA)**
  - Offline one-worker-many-tasks assignment

  - One-worker-one-task → **Minimum Bipartite Matching problem (MBM)**
  - One-worker-many-tasks → **Extended Minimum Bipartite Matching problem (EMBM)**
4. Approaches

- Threshold-based Online task Assignment mechanism (TOA)
  - Offline one-worker-many-tasks assignment

\[\text{Algorithm 2: Extended minimum-cost flow (EMCF)}\]

\begin{algorithm}
\caption{Extended minimum-cost flow (EMCF)}
\begin{algorithmic}
    \Statex \textbf{Input:} Workers $\mathcal{W}$, tasks $\mathcal{T}$, cardinality constraint $\delta$
    \Statex \textbf{Output:} The minimum-cost flow $\mathcal{F}$, total cost $\mathcal{D}$
    \State Construct $G' = (V', A')$ as Eq. (10) based on $\mathcal{W}, \mathcal{T}$
    \State Initialize the flow $\mathcal{F} \leftarrow \emptyset$, total cost $\mathcal{D} \leftarrow 0$
    \For{$(w_i, t_j) \in A'$}
        \If{$w_i.t_l < t_j.t_a$ or $w_i.t_a > t_j.t_l$}
            \State Remove the arc $(w_i, t_j)$ from $A'$;
        \EndIf
    \EndFor
    \State Find the minimum-cost augmenting path $\mathcal{P}(G', s, t)$
    \While{$\mathcal{P}(G', s, t)$ exists and $|\mathcal{F}| < \delta$}
        \State $\mathcal{F} \leftarrow \mathcal{F} \cup \mathcal{P}(G', s, t)$
        \For{$\text{arc} (v', v) \in \mathcal{P}(G', s, t)$}
            \If{$\text{arc} (v', v) \notin A'$}
                \State $A' \leftarrow A' \cup \{(v', v)\}$
                \State $d(v', v) \leftarrow -d(v', v')$, $c(v', v) \leftarrow 0$
                \State $c(v, v') \leftarrow c(v, v') - 1$, $c(v', v) \leftarrow c(v', v) + 1$
                \State $\mathcal{D} \leftarrow \mathcal{D} + d(v, v')$
                \If{$v' = t_j, \forall t_j \in \mathcal{T}$}
                    \State $d(s, t'_j) \leftarrow 0$
                \EndIf
            \EndIf
        \EndFor
        \State Find $\mathcal{P}(G', s, t)$ again based on the current $G'$
    \EndWhile
    \State \Return $\mathcal{F}$, $\mathcal{D}$
\end{algorithmic}
\end{algorithm}
4. Approaches

- Threshold-based Online task Assignment mechanism (TOA)
  - Online one-worker-many-tasks assignment

✓ TOA estimates a threshold

$$\kappa = \min\{d(\mathcal{P})\}, \forall \mathcal{P} \in \mathcal{F}$$

✓ Theorem 2. TOA achieves a bound of

$$\frac{\phi \cdot \delta}{d(O') - \sigma}$$

$$\Pr[|d(O) - E(d(O'))| \leq \varepsilon] \geq 1 - \sigma^2 / \varepsilon^2,$$

$$CR = E(TOA)/OPT = E(\delta \cdot \Delta) / d(O) = \delta E(\Delta) / d(O) \leq \phi \cdot \delta / d(O) \leq \phi \cdot \delta / (1 - \sigma^2 / \varepsilon^2) \cdot (d(O') - \varepsilon). \quad (13)$$

✓ Theorem 3. PLP-TOA achieves a bound of $O(1/(\bar{d} \cdot \varepsilon^2))$

$$\Pr[|d(\hat{\mathcal{F}}) - d(\mathcal{F})| \geq \lambda] \geq \frac{6}{\lambda^2 \cdot \varepsilon^2}, \quad CR = d(\hat{\mathcal{F}})/OPT \leq \frac{6(d(\mathcal{F}) + \lambda)}{d(O) \cdot \lambda^2 \varepsilon^2} \leq \frac{6(\kappa \cdot \delta + \lambda)}{\lambda^2 \varepsilon^2 (d(O') - \sigma)},$$

Algorithm 3: Threshold online assignment (TOA)

```
Input: \mathcal{W}, \mathcal{T}, \delta, T, \phi
Output: \mathcal{F}
1 \mathcal{F} \leftarrow \emptyset, t \leftarrow 0 ;
2 \textbf{while} t \leq T \textbf{and} |\mathcal{F}| < \delta \textbf{do}
3 \quad \mathcal{W}_t, \mathcal{T}_t \leftarrow \text{the current workers and tasks in } t;
4 \quad \mathcal{F}_t, \mathcal{D}_t = \text{EMCF}(\mathcal{W}_t, \mathcal{T}_t, \delta);
5 \quad \textbf{foreach} \mathcal{P} \in \mathcal{F}_t \textbf{do}
6 \quad \quad \textbf{if } d(\mathcal{P}) \leq \kappa \text{ \textbf{and} the pre-path of } \mathcal{P} \text{ \textbf{is in } } \mathcal{F} \textbf{then}
7 \quad \quad \quad \mathcal{F} \leftarrow \mathcal{F} \cup \{\mathcal{P}\};
8 \quad \textbf{foreach} \mathcal{P} \in \mathcal{F}_t \textbf{do}
9 \quad \quad \quad \textbf{return } \mathcal{F}
```
5. Experiments

1. Introduction
2. Challenges and contributions
3. Problem formulation
4. Approaches
5. Experiments
5. Experiments

- Two real-world datasets, three cities
  - Gowalla: New York
  - Foursquare: Tokyo, London

- Baselines
  - PLP-TOA
  - TBF (ICDE’20 [1])
  - PLP-OA
  - OPT
  - PLP-Gre

- Settings

5. Experiments

RQ1: Does our method deal with the online one-worker-many-tasks assignment?

Fig. 14: Examples of the one-worker-many-tasks assignment.

Fig. 13: Total distance vs. Capacity.
5. Experiments

RQ2: Does our method agree with the theoretical results?

PLP-TOA achieves a bound of $O(1/(\bar{d} \cdot \epsilon^2))$

$$CR = d(\tilde{F})/OPT \leq \frac{6(d(F) + \lambda)}{d(O) \cdot \lambda^2 \epsilon^2} \leq \frac{6(\kappa \cdot \delta + \lambda)}{\lambda^2 \epsilon^2(d(O') - \sigma)},$$

Fig. 12: Competitive ratio.

(a) Tokyo  (b) New York  (c) London
5. Experiments

- **RQ3**: How different parameters affect simulation results?

![Graphs showing total distance vs. worker number and task number for Tokyo, New York, and London.](graphs)

**Fig. 8**: Total distance vs. Worker number.

**Fig. 9**: Total distance vs. Task number.
5. Experiments

RQ3: How different parameters influence our mechanism?

Fig. 10: Total distance vs. Privacy budget.

Fig. 11: Total distance vs. Cardinality.
Thank you!