Abstract

Recently, it becomes easy to collect our own data in real life such as GPS locations. Can we utilize personal data without privacy violation? In this work, we proposed:

- $\ell$-trajectory privacy model providing personalized privacy control, and
- an algorithm framework to achieve this privacy goal, meanwhile keeping high data utility.

Introduction

A $\ell$-trajectory: $\ell$ successive accessed locations by a user.

<table>
<thead>
<tr>
<th>t1</th>
<th>t2</th>
<th>t3</th>
<th>t4</th>
<th>t5</th>
</tr>
</thead>
<tbody>
<tr>
<td>u1</td>
<td>park</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>u2</td>
<td>bar</td>
<td>office</td>
<td>gym</td>
<td></td>
</tr>
<tr>
<td>u3</td>
<td>gym</td>
<td>office</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t1</th>
<th>t2</th>
<th>t3</th>
<th>t4</th>
<th>t5</th>
</tr>
</thead>
<tbody>
<tr>
<td>park</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>office</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>gym</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Example: Qcount: How many people at Kyoto station now?

Model

$\varepsilon$-Differential Privacy ($\varepsilon$-DP) is a de facto model for Privacy Preserving Data Publishing (PPDP). Its idea is to cover sensitive data by adding random noises to satisfy

$$\Pr(Q(D)) \leq e^\varepsilon \cdot \Pr(Q(D^\circ)), \quad \varepsilon > 0$$

$\ell$-trajectory privacy is based on $\varepsilon$-DP:
- to limit the impact of any single $\ell$-trajectory to the query result
- to make sure any $\ell$-trajectory under $\varepsilon$-DP
- User-level Protection (v.s. Event-level [3])

Experiments

Experiments on four real-life datasets show that GA+MMD has relatively high data utility. FAST is a competitor PPDP framework for time-series data. Uniform is a baseline method of uniform budget allocation to satisfy $\ell$-trajectory privacy.

GA+MMD's noisy data is closer to the real data. MAE: (Mean of Absolute Error) the lower the less noise. KL-Divergence: the lower the similar to original data's distribution.

Algorithms

Proposed Framework of PPDP over Infinite Trajectory Streams

- For Infinite Streams (v.s. Finite streams)
- Real-time (v.s. offline publishing)

To Save Privacy budget (Adj or MMD)

To integrate Information of other components

Dynamic Budget Allocation

Private Approximation Strategy

Private Publishing

A previous model: w-event privacy [3] (w=3)

Theorem 2: (feasibility of $\ell$-trajectory privacy) If the sum of privacy budgets on any $\ell$-trajectory is less than or equal to $\varepsilon$, we can obtain $\ell$-trajectory privacy.

Kyoto University, Department of Social Informatics, Yoshikawa-Ma Lab.