Differentially Private Correlation Heatmap from Multi-Modal Location Datasets

Sijie Xiong, Hafiz Imtiaz (advisors: Prof. D. Zhang, Prof. A. D. Sarwate)

Rutgers, The State University of New Jersey

Motivation

Goal: find "Points of Interest" in a city
→ can use *location entropy* [1]
Challenge: location data are private and sparse
→ need to preserve privacy
→ can use multi-modal datasets
Can we find a better approach?

Problem Setup

Raw dataset snippet (`taxi.txt` and `bus.txt`)

bus.txt

```
ttAqkYWPav8=,2014-03-03T07:26:41.000Z,114.110115,22.543484,M3733
ttAqkYWPav8=,2014-03-03T14:24:01.000Z,114.110214,22.542334,M3733
ttAqkYWPav8=,2014-03-03T14:24:21.000Z,114.110237,22.542334,M3733
ttAqkYWPav8=,2014-03-03T20:27:46.000Z,114.110451,22.5415,M3733
ttAqkYWPav8=,2014-03-03T07:33:21.000Z,114.110481,22.540318,M3733
```

- Recordings are from a regular Monday (12h period)
- Divide entire city into 200×1000-grid $\mathbb{G}$
- Compute location entropy for all locations $l \in \mathbb{G}$ for taxis ($\mathbf{X} \in \mathbb{R}^{D_x \times N}$) and buses ($\mathbf{Y} \in \mathbb{R}^{D_y \times N}$)

Location Entropy [2]

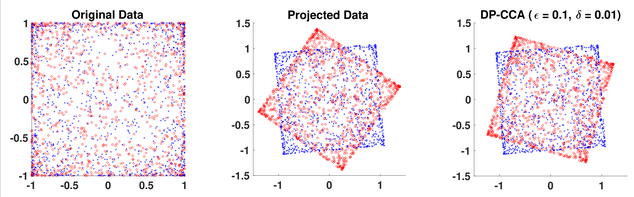
Given a location $l \in \mathbb{G}$,

- $\mathbb{S}_l$, the set of visits to l
- $\mathbb{S}_{l,v}$, the set of visits vehicle v has made to l
- $p_{l,v} = |\mathbb{S}_{l,v}|/|\mathbb{S}_l|$, the fraction of total visits to l that belongs to vehicle v
- $\mathbb{V}_l$, the set of unique vehicles that visited l

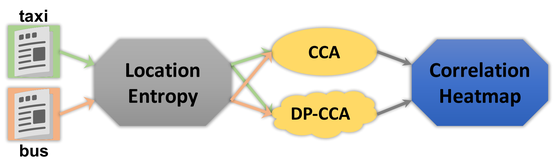
Location Entropy: $H(l) = -\sum_{v \in \mathbb{V}_l} p_{l,v} \log p_{l,v}$
→ measures both the *frequency* and *diversity* of visits.

Canonical Correlation Analysis (CCA)

CCA finds subspaces for different modes of data
→ modes are maximally correlated after projection



Processing Pipeline

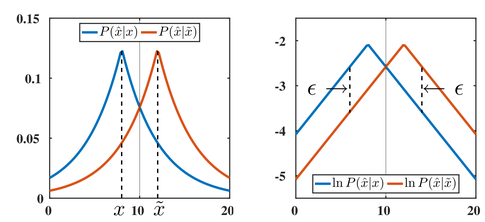


Differential Privacy (DP)

DP is *formal* and *quantifiable*.
Definition: Algorithm $\mathcal{A}(\mathbb{D})$ taking values in a set $\mathbb{T}$ provides (ϵ, δ)-differential privacy if

$$P(\mathcal{A}(\mathbb{D}) \in \mathbb{S}) \leq e^{\epsilon} P(\mathcal{A}(\mathbb{D}') \in \mathbb{S}) + \delta,$$

for all measurable $\mathbb{S} \subseteq \mathbb{T}$ and all *neighboring* data sets $\mathbb{D}$ and $\mathbb{D}'$ differing in a single entry.
Interpretation: $(\epsilon, \delta) \downarrow \Rightarrow$ privacy level $\uparrow$



DP-CCA [3]

DP-CCA adds noise to the covariance matrix.
Algorithm:

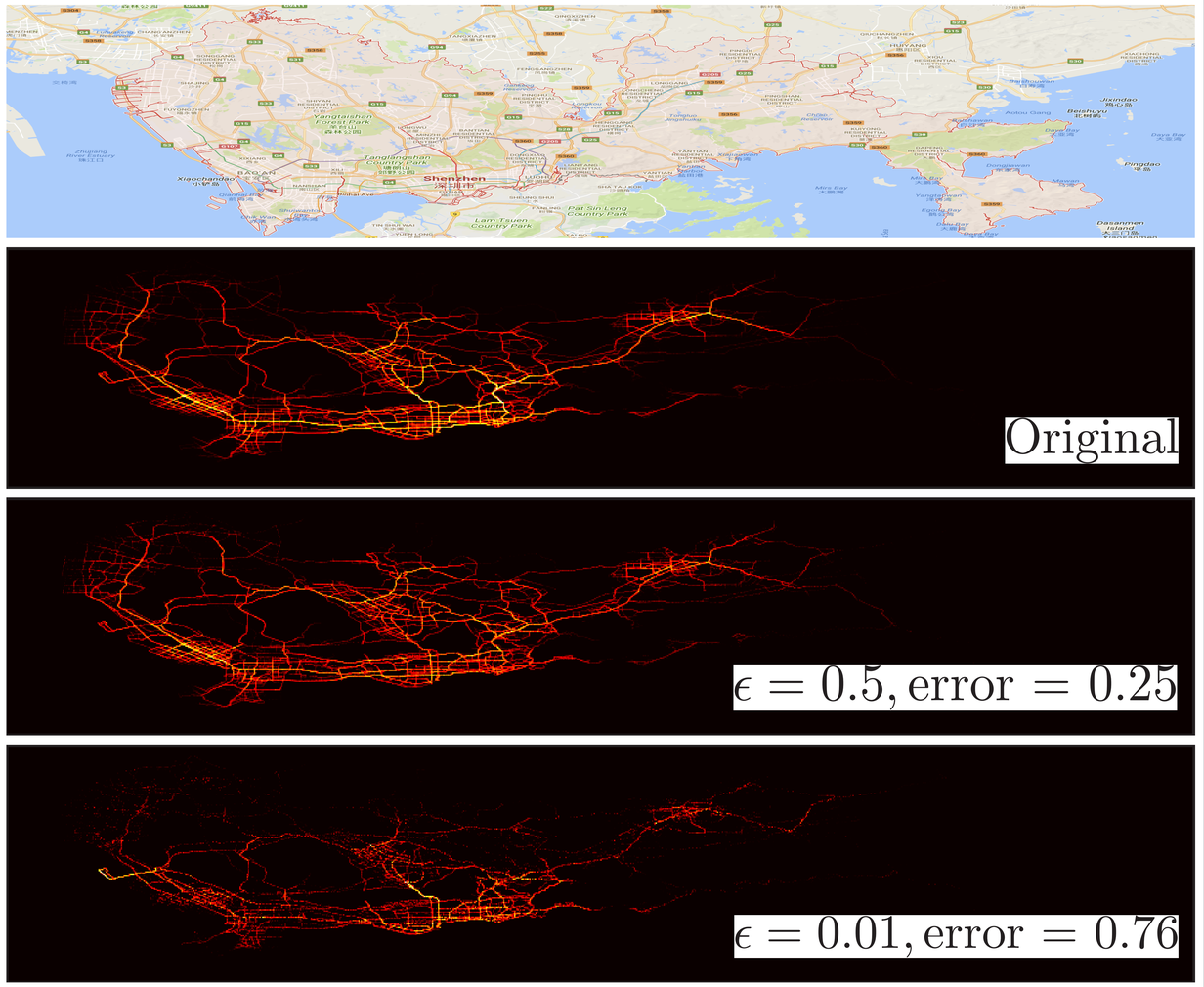
- Obtain mean-centered and normalized data $\mathbf{Z} = [\mathbf{X}; \mathbf{Y}]$
- Compute $\mathbf{C} = \frac{1}{N}\mathbf{Z}\mathbf{Z}^\top$
- Compute $\hat{\mathbf{C}} = \mathbf{C} + \mathbf{E}$, where $\mathbf{E}$ is a noise matrix calibrated to satisfy DP

$$\hat{\mathbf{C}} = \begin{bmatrix} \hat{\mathbf{C}}_{xx} & \hat{\mathbf{C}}_{xy} \\ \hat{\mathbf{C}}_{xy}^\top & \hat{\mathbf{C}}_{yy} \end{bmatrix}$$

- $\mathbf{U}$: top-K eigenvectors of $\hat{\mathbf{C}}_{xx}^{-1}\hat{\mathbf{C}}_{xy}\hat{\mathbf{C}}_{yy}^{-1}\hat{\mathbf{C}}_{yx}$
- $\mathbf{V}$: top-K eigenvectors of $\hat{\mathbf{C}}_{yy}^{-1}\hat{\mathbf{C}}_{yx}\hat{\mathbf{C}}_{xx}^{-1}\hat{\mathbf{C}}_{xy}$

Correlation Heatmap:
→ visualize $\mathrm{tr}\left[(\mathbf{U}^\top \mathbf{x}_n)^\top (\mathbf{V}^\top \mathbf{y}_n)\right], \forall n \in [N]$

Simulation Results



Remark: as $\epsilon \downarrow$, the DP-CCA heatmap becomes noisier, but still offers good estimation.

Future Work

- Correlation of traffic speeds among different modes of transportation
- Use the correlation to improve navigation

References

[1] C. Justin et al. "Bridging the gap between physical location and online social networks," in Proceedings of the 12th ACM international conference on Ubiquitous computing. ACM, 2010.
[2] T. Hien et al. "Differentially private publication of location entropy," in Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, 2016.
[3] H. Imtiaz and A. D. Sarwate. "Differentially private canonical correlation analysis," in Proceedings of the 2017 IEEE Global Conference on Signal and Information Processing. GlobalSIP, 2017, to appear.