Traffic Monitoring using Differential Privacy

Shubham Aggarwal, Dionysis Manousakas

University of Cambridge

Extended Abstract

User location data is being widely used to drive many modern software systems: apps like Waze and Uber heavily rely on such data to function, systems use publicly available location datasets for training etc. However, this raises some natural privacy concerns. Services collecting such data can deduce user habits and behaviour and even anonymised public data might leak some private information[1]. In this work we explore and evaluate the statistical technique of Differential Privacy as a way of addressing such issues.

Differential Privacy provides a mathematical framework that allows us to rigorously quantify individual privacy in a dataset and preserve it across any future transformation to the data. This privacy measure is achieved by introducing carefully generated random noise into the dataset; but naturally, this makes the dataset less accurate. The accuracy loss can be quantitatively measured to compare the performance of different methods within this framework.

As differentially private data is robust against any post-processing, we can apply some denoising techniques which exploit properties inherent to location data to recover some of the accuracy lost. We implement a few such methods and compare the improvement in accuracy each of them provides to privatised datasets generated by different differential privacy mechanisms.

Finally, we analyse how using such privatisation techniques affects performance in a real-world application. We evaluate the performance of an algorithm that solves the k-taxi dispatch problem: at what locations should a fleet of k taxis be placed to best cover the demand for taxi rides in a city. We compare how the performance of this algorithm is affected when using private datasets generated using different combinations of differential privacy mechanisms and post-processing techniques as opposed to the non-private dataset.

Based on our experiments, we conclude that using targeted post-processing techniques in conjunction with differential privacy can allow for the design of services that rely on user location data while maintaining individual-level privacy.

[1] `https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/`