Data in Brief 20 (2018) 1510–1514

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Differentially private count queries over personalized-location trajectory databases

Fatemeh Deldar, Mahdi Abadi *

Faculty of Electrical and Computer Engineering, Tarbiat Modares University, P.O. Box 14115-194, Tehran, Iran

ARTICLE INFO

Article history:
Received 16 June 2018
Accepted 24 August 2018
Available online 3 September 2018

Keywords:
Differential privacy
Count query
Personalized-location trajectory dataset
Benchmark dataset

ABSTRACT

Differential privacy is a technique for releasing statistical information about a database without revealing information about its individual data records. Also, a personalized-location trajectory database is a trajectory database where locations have different privacy protection requirements and, thus, are privacy conscious. This data article is related to the research article entitled "PLDP-TD: Personalized-location differentially private data analysis on trajectory databases" (Deldar and Abadi, 2018 [1]), in which we introduced a new differential privacy notion for personalized-location trajectory databases, and devised a novel differentially private algorithm, called PLDP-TD, to implement this new privacy notion. Here, we describe how the datasets in the research article were obtained and measure the relative error of PLDP-TD for different non-zero count query sets.

Specifications table

Subject area	Computer science
More specific subject area	Information security; Data privacy; Differential privacy
Type of data	Table, figure, text file
How data was acquired	Processing of synthetic and real trajectory data

DOI of original article: http://dx.doi.org/10.1016/j.pmcj.2018.06.005
* Corresponding author. Fax: +98 21 82884325.
E-mail addresses: f.deldar@modares.ac.ir (F. Deldar), abadi@modares.ac.ir (M. Abadi).

http://dx.doi.org/10.1016/j.dib.2018.08.104

Data format	Raw, analyzed
Experimental factors	The continuous spatial domain was discretized by partitioning it into a uniform grid and then each grid cell was considered as a location. Then, the latitude-longitude coordinates in each input trajectory dataset were replaced by their corresponding locations. Afterwards, a privacy descriptor was assigned to each location, resulting in a personalized-location trajectory dataset.
Experimental features	A differentially private algorithm was applied to each personalized-location trajectory dataset and the average relative error of noisy answers to non-zero count queries was measured.
Data source location	N/A
Data accessibility	Data are with this article.

Value of the data

- The obtained personalized-location trajectory datasets can be used as benchmark datasets to examine the effectiveness of various differentially private algorithms developed for trajectory databases.
- The method used for obtaining personalized-location trajectory datasets can be used to obtain personalized datasets in other contexts.
- The personal privacy budget allocation strategy can be used as a reference strategy for effective non-uniform budget allocation in any tree-based structure.
- The results showed that by providing non-uniform privacy guarantees, the quality of noisy answers to count queries significantly improves.

1. Data

Our personalized-location trajectory datasets are obtained based on three synthetic and real trajectory datasets, namely, City80K [2], Geolife [3], and Taxi [4]. The City80K dataset simulates the routes of 80,000 citizens in a metropolitan area with 26 city blocks in 24 hours. The Geolife dataset collects the GPS trajectories of 182 users in Beijing, China, during a period of over five years (from April 2007 to August 2012), including not only common activities like go home and go to work, but also some entertainments and sports activities, such as shopping, hiking, and cycling. We choose the trajectories whose latitude-longitude coordinates (moving points) are between [39.4, 40.8] latitude and [115.8, 117.4] longitude, and break a trajectory if the interval time between its subsequent and current latitude-longitude coordinates is larger than one minute. Since each recorded trajectory consists of a sequence of latitude-longitude coordinates in a continuous spatial domain, we discretize the spatial domain by partitioning it into a 64×64 grid and then consider each grid cell as a location. We further replace the latitude-longitude coordinates by the obtained locations. Finally, the Taxi dataset contains approximately 5.2 million GPS trajectories of 8602 taxi cabs in Beijing, China, recorded during a 1-month period in May 2009. The trajectory data cover a region of Beijing

Table 1
Summary of trajectory datasets.

Dataset	No. of trajectories	Type of spatial domain	No. of locations	Average trajectory length	Maximum trajectory length
City80K	80,000	Discrete	26	4.52	24
Geolife	1,028,434	Continuous	4096	19.52	601
Taxi	5,251,604	Continuous	256	11.32	53,095