Title

Optimizing Linear Queries Under Differential Privacy

Author

Chao Li, *University of Massachusetts Amherst* Follow

Date of Award

9-2013

Document Type

Open Access Dissertation

Degree Name

Doctor of Philosophy (PhD)

Degree Program

Computer Science

First Advisor

Gerome Miklau

Second Advisor

Andrew McGregor

Third Advisor

Don Towsley

Subject Categories

Computer Sciences

Abstract

Private data analysis on statistical data has been addressed by many recent literatures. The goal of such analysis is to measure statistical properties of a database without revealing information of individuals who participate in the database. Differential privacy is a rigorous privacy definition that protects individual information using output perturbation: a differentially private algorithm produces statistically indistinguishable outputs no matter whether the database contains a tuple corresponding to an individual or not.

It is straightforward to construct differentially private algorithms for many common tasks and there are published algorithms to support various tasks under differential privacy. However methods to design error-optimal algorithms for most non-trivial tasks are still unknown. In particular, we are interested in error-optimal algorithms for sets of linear queries. A linear query is a sum of counts of tuples that satisfy a certain condition, which covers the scope of many aggregation tasks including count, sum and histogram. We present the matrix mechanism, a novel mechanism for answering sets of linear queries under differential privacy. The matrix mechanism makes a clear distinction between a set of queries submitted by users, called the query workload, and an alternative set of queries to be answered under differential privacy, called the query strategy. The answer to the query workload can then be computed using the answer to the query strategy. Given a query workload, the query strategy determines the distribution of the output noise and the power of the matrix mechanism comes from adaptively choosing a query strategy that minimizes the output noise.

Our analyses also provide a theoretical measure to the quality of different strategies for a given workload. This measure is then used in accurate and approximate formulations to the optimization problem that outputs the error-optimal strategy. We present a lower bound of error to answer each workload under the matrix mechanism. The bound reveals that the hardness of a query workload is related to the spectral properties of the workload when it is represented in matrix form. In addition, we design an approximate algorithm, which generates strategies generated by our a out perform state-of-art mechanisms over (epsilon, delta)-differential privacy. Those strategies lead to more accurate data analysis while preserving a rigorous privacy guarantee. Moreover, we also combine the matrix mechanism with a novel data-dependent algorithm, which achieves differential privacy by adding noise that is adapted to the input data and to the given query workload.

Recommended Citation

Li, Chao, "Optimizing Linear Queries Under Differential Privacy" (2013). *Open Access Dissertations*. 805.
https://scholarworks.umass.edu/open_access_dissertations/805

Download

DOWNLOADS

Since November 26, 2013