Title:

Generating Synthetic Data based on Noise Multiplied Data

Abstract:

With or without particular purposes, much data is collected by government agencies (like Census Bureaus and National Statistical Offices) and Internet media (such as Google, Facebook and Twitter). Making this data available for commercial and public use brings many benefits to the users and data providers. Because the collected data might be confidential, it is important that confidentiality of the data not be compromised when it is used in the public sphere. As a result a certain level of protection is essential in order to maintain the privacy of the individuals in question.

Synthetic data that mimic the original data and preserve the relationships between variables is a possible solution to data privacy. Currently, synthetic data are generated by data vendors using model-based approaches, which is inconvenient to both data vendors and users. The data users have no control over the models from which synthetic data were generated.

This talk proposes a framework of a new approach to provide a viable alternative to existing approaches. By following the new approach, "synthetic data" are produced by the data user instead of the data provider. No model is required for generating the "synthetic data". The data user might have more freedom in the process of data exploration by following the approach proposed compared to existing approaches.