Compressive Privacy Based on Joint Optimization of Differential Utility/Cost

Keynote Speech by S.Y. Kung

IEEE Conference on Big Data Security on Cloud, 2016

With the rapidly growing internet commerce, personal data are being collected, stored, and circulated around the internet - often without the data owner's knowledge. As such, protection of *privacy of internet data* has become a vital research field. Conventionally, the task of data protection is left entirely to the cloud server, rendering data privacy extremely vulnerable to hacker attack or unauthorized leakage. To rectify this problem, a key is to let data owners (not cloud servers) control the data privacy.

This talk explores the rich synergy among signal processing, information theory, estimation theory, and machine learning and, thereafter, presents a novel privacy preserving methodology, named *compressive privacy* (CS). The objective is to build user-collaborative machine learning systems to facilitate the intended function while protecting the privacy of owner's sensitive information. It involves the joint optimization over three design spaces: (i) Feature Space (ii) Utility Subspace; and (iii) Cost Subspace (i.e. Privacy Subspace).

Mathematically, the optimal query can be derived from the joint optimization formulation where the query should be chosen to simultaneously maximize the utility and minimize the cost. In order to derive a closed form analysis/solution, we recast the information theoretical criterion (such as the log-likelihood or mutual information) in terms of (differential) *error covariance matrix* used in the estimation theory. More exactly, the optimal query search (or feature selection) becomes a problem of maximizing a *Differential Utility/Cost (DUC)* ratio, a criterion function commonly adopted by economists. More exactly, DUC is defined as the ratio between *Differential Utility* and *Differential Cost*.

The DUC formulation can be extended to Machine Learning applications, where the *Differential Utility* and *Cost* are characterized by the given supervised training dataset.
Furthermore, the DUC optimization can be reformulated in the kernel learning models, where nonlinear kernels afford a much expanded search space to enhance the optimized DUC ratio.

Simulation results based on facial image classification (utility) and reconstruction (privacy) will be demonstrated.

Acknowledgement: This material is based upon work supported in part by the *Brandeis Program* of the Defense Advanced Research Project Agency (DARPA) and Space and Naval Warfare System Center Pacific (SSC Pacific) under Contract No. 66001-15-C-4068.