Josep Domingo Ferrer

TITLE: Information-theoretic risk and utility measures for microdata

ABSTRACT: Information loss measures based on relative discrepancies of statistics between original and masked/synthetic data are awkward to combine with disclosure risk measures. The reason is that relative discrepancies are unbounded (at least for continuous attributes). Probabilistic information loss measures are a first attempt at offering bounded information loss measures, but they lack a theoretical framework which would be of great help to optimize the trade-off between loss and risk. We discuss here information-theoretic measures for information loss and disclosure risk. These are heavily based on mutual information. We shown how to express data perturbation and synthetic data generation as optimization problems involving the proposed measures.