Project ideas

Shumin Guo

1. Implement the data reconstruction algorithms and repeat the experiments in limiting privacy breaches in privacy preserving data mining, i.e., paper 15. Then, if we change the distribution assumptions on original data and noise, how effective are the different methods, particularly the Bayes method?

   References: 13,15

2. Compare additive perturbation, random projection perturbation, and geometric data perturbation, on a few experimental datasets. Investigate the relationship between the loss of privacy and the model accuracy. Will random projection perturbation and geometric data perturbation preserve the accuracy of decision tree classifier? Will the additive perturbation method preserve the accuracy of distance-based classifiers, such as k-nearest-neighbor classifier?

   References: 10,20,21

3. Compare l-diversity, t-closeness and k-anonymity, and study how data utility is sacrificed in each model (use decision tree for experiments)

   References: 39,53

4. Simulate graph structures (paper 182), and implement the attacks on graph structure in paper 167.

   References: 167, 182

5. Study the social network data publishing problem. Pick one of the papers 163-166

   References: 163-166

6. Study the location privacy problem, and implement algorithms and repeat some experiments in the paper 175

   References: 170,173,174, 175

7. Study the high-dimensional data perturbation problem

   References: 29,51,64

8. Consider secure outsourced data for k-nearest neighbor search services.

   References: 187,189,190,191