

Protecting Privacy Against Classification Attacks in Data Mining

Xiaobai (Bob) Li
College of Management
UMass Lowell

Wednesday, 22 February 2006

Olsen 311

Refreshments at 2:30, Talk from 3:00-4:00

Data mining techniques can be used not only to study collective behavior about customers, but also to discover private information about individuals. We demonstrate that classification trees, a popular data mining technique, can be used to effectively reveal individuals confidential data, even when the identities of the individuals are not present in the data. We propose a method for organizations to protect confidential data from such a classification attack. The proposed method adopts an over-pruning strategy to classification trees to identify records with high disclosure risks, and applies a data swapping procedure to reduce the disclosure risks while preserving data quality. An experimental study on two real-world datasets shows that the proposed method is very effective for privacy-preserving data mining.

Bio: Dr. Li is an Assistant Professor of Management Information Systems at UMass Lowell. He received his Ph.D. in management science from University of South Carolina. His research interests include data mining, data privacy, and decision support systems. He has published in IEEE Transactions on Automatic Control, IEEE Transactions on Systems, Man, and Cybernetics, INFORMS Journal on Computing, Decision Support Systems, European Journal of Operational Research, among others.