Data Privacy

Rakesh Agrawal

IBM Almaden Research Center, San Jose, CA 95120, USA

There is increasing need to build information systems that protect the privacy and ownership of data without impeding the flow of information. We will present some of our current work to demonstrate the technical feasibility of building such systems:

Privacy-preserving data mining. The conventional wisdom held that data mining and privacy were adversaries, and the use of data mining must be restricted to protect privacy. Privacy-preserving data mining cleverly finesses this conflict by exploting the difference between the level where we care about privacy, i.e., individual data, and the level where we run data mining algorithms, i.e., aggregated data. User data is randomized such that it is impossible to recover anything meaningful at the individual level, while still allowing the data mining algorithms to recover aggregate information, build mining models, and provide actionable insights.

Hippocratic databases. Unlike the current systems, Hippocratic databases include responsibility for the privacy of data they manage as a founding tenet. Their core capabilities have been distilled from the principles behind current privacy legislations and guidelines. We identify the technical challenges and problems in designing Hippocratic databases, and also outline some solutions.

Sovereign information sharing. Current information integration approaches are based on the assumption that the data in each database can be revealed completely to the other databases. Trends such as end-to-end integration, outsourcing, and security are creating the need for integrating information across autonomous entities. In such cases, the enterprises do not wish to completely reveal their data. In fact, they would like to reveal minimal information apart from the answer to the query. We have formalized the problem, identified key operations, and designed algorithms for these operations, thereby enabling a new class of applications, including information exchange between security agencies, intellectual property licensing, crime prevention, and medical research.

References

1. R. Agrawal, R. Srikant: Privacy Preserving Data Mining. ACM Int'l Conf. on Management of Data (SIGMOD), Dallas, Texas, May 2000.
2. R. Agrawal, J. Kiernan, R. Srikant, Y. Xu: Hippocratic Databases. 28th Int'l Conf. on Very Large Data Bases (VLDB), Hong Kong, August 2002.
3. R. Agrawal, A. Evfimievski, R. Srikant: Information Sharing Across Private Databases. ACM Int'l Conf. on Management of Data (SIGMOD), San Diego, California, June 2003.

J.-F. Boulicaut et al. (Eds.): PKDD 2004, LNAI 3202, p. 8, 2004.