Privacy-Preserving Data Publishing

An Overview

Synthesis Lectures on Data Management

Editor

M. Tamer Özsu, *University of Waterloo*

Synthesis Lectures on Data Management is edited by Tamer Özsu of the University of Waterloo. The series will publish 50- to 125 page publications on topics pertaining to data management. The scope will largely follow the purview of premier information and computer science conferences, such as ACM SIGMOD, VLDB, ICDE, PODS, ICDT, and ACM KDD. Potential topics include, but not are limited to: query languages, database system architectures, transaction management, data warehousing, XML and databases, data stream systems, wide scale data distribution, multimedia data management, data mining, and related subjects.

Privacy-Preserving Data Publishing: An Overview

Raymond Chi-Wing Wong and Ada Wai-Chee Fu
2010

An Introduction to Duplicate Detection

Felix Naumann and Melanie Herschel
2010

Keyword Search in Databases

Jeffrey Xu Yu, Lu Qin, and Lijun Chang
2010

Privacy-Preserving Data Publishing: An Overview
Raymond Chi-Wing Wong and Ada Wai-Chee Fu

ISBN: 978-3-031-00706-4 paperback
ISBN: 978-3-031-01834-3 ebook

DOI 10.1007/978-3-031-01834-3

A Publication in the Springer series
SYNTHESIS LECTURES ON DATA MANAGEMENT

Lecture #3
Series Editor: M. Tamer Özsu, *University of Waterloo*
Series ISSN
Synthesis Lectures on Data Management
ISSN pending.

Privacy-Preserving Data Publishing

An Overview

Raymond Chi-Wing Wong
The Hong Kong University of Science and Tecnology

Ada Wai-Chee Fu
The Chinese University of Hong Kong

SYNTHESIS LECTURES ON DATA MANAGEMENT #3

ABSTRACT

Privacy preservation has become a major issue in many data analysis applications. When a data set is released to other parties for data analysis, privacy-preserving techniques are often required to reduce the possibility of identifying sensitive information about individuals. For example, in medical data, sensitive information can be the fact that a particular patient suffers from HIV. In spatial data, sensitive information can be a specific location of an individual. In web surfing data, the information that a user browses certain websites may be considered sensitive. Consider a dataset containing some sensitive information is to be released to the public. In order to protect sensitive information, the simplest solution is not to disclose the information. However, this would be an overkill since it will hinder the process of data analysis over the data from which we can find interesting patterns. Moreover, in some applications, the data must be disclosed under the government regulations. Alternatively, the data owner can first modify the data such that the modified data can guarantee privacy and, at the same time, the modified data retains sufficient utility and can be released to other parties safely. This process is usually called as privacy-preserving data publishing. In this monograph, we study how the data owner can modify the data and how the modified data can preserve privacy and protect sensitive information.

KEYWORDS

privacy preservation, data publishing, anonymity, data mining

Contents