Privacy-Preserving Data Mining

Models and Algorithms

ADVANCES IN DATABASE SYSTEMS

Volume 34

Series Editors

Ahmed K. Elmagarmid
Purdue University
West Lafayette, IN 47907

Amit P. Sheth
Wright State University
Dayton, Ohio 45435

Other books in the Series:

SEQUENCE DATA MINING, Guozhu Dong, Jian Pei; ISBN: 978-0-387-69936-3
DATA STREAMS: *Models and Algorithms*, edited by Charu C. Aggarwal; ISBN: 978-0-387-28759-1
SIMILARITY SEARCH: ***The Metric Space Approach***, *P. Zezula, G. Amato, V. Dohnal, M. Batko*; ISBN: 0-387-29146-6
STREAM DATA MANAGEMENT, *Nauman Chaudhry, Kevin Shaw, Mahdi Abdelguerfi*; ISBN: 0-387-24393-3
FUZZY DATABASE MODELING WITH XML, *Zongmin Ma*; ISBN: 0-387- 24248-1
MINING SEQUENTIAL PATTERNS FROM LARGE DATA SETS, *Wei Wang and Jiong Yang*; ISBN: 0-387-24246-5
ADVANCED SIGNATURE INDEXING FOR MULTIMEDIA AND WEB APPLICATIONS, *Yannis Manolopoulos, Alexandros Nanopoulos, Eleni Tousidou*; ISBN: 1-4020-7425-5
ADVANCES IN DIGITAL GOVERNMENT: *Technology, Human Factors, and Policy*, *edited by William J. McIver, Jr. and Ahmed K. Elmagarmid*; ISBN: 1-4020-7067-5
INFORMATION AND DATABASE QUALITY, *Mario Piattini, Coral Calero and Marcela Genero*; ISBN: 0-7923-7599-8
DATA QUALITY, *Richard Y. Wang, Mostapha Ziad, Yang W. Lee:* ISBN: 0-7923-7215-8
THE FRACTAL STRUCTURE OF DATA REFERENCE: *Applications to the Memory Hierarchy*, *Bruce McNutt*; ISBN: 0-7923-7945-4
SEMANTIC MODELS FOR MULTIMEDIA DATABASE SEARCHING AND BROWSING, *Shu-Ching Chen, R.L. Kashyap, and Arif Ghafoor*; ISBN: 0-7923-7888-1
INFORMATION BROKERING ACROSS HETEROGENEOUS DIGITAL DATA: A Metadata-based Approach, *Vipul Kashyap, Amit Sheth*; ISBN: 0-7923-7883-0
DATA DISSEMINATION IN WIRELESS COMPUTING ENVIRONMENTS, *Kian-Lee Tan and Beng Chin Ooi*; ISBN: 0-7923-7866-0
MIDDLEWARE NETWORKS: Concept, Design and Deployment of Internet Infrastructure, *Michah Lerner, George Vanecek, Nino Vidovic, Dad Vrsalovic*; ISBN: 0-7923-7840-7
ADVANCED DATABASE INDEXING, *Yannis Manolopoulos, Yannis Theodoridis, Vassilis J. Tsotras*; ISBN: 0-7923-7716-8
MULTILEVEL SECURE TRANSACTION PROCESSING, *Vijay Atluri, Sushil Jajodia, Binto George* ISBN: 0-7923-7702-8
FUZZY LOGIC IN DATA MODELING, *Guoqing Chen* ISBN: 0-7923-8253-6
PRIVACY-PRESERVING DATA MINING: Models and Algorithms, *edited by Charu C. Aggarwal and Philip S. Yu*; ISBN: 0-387-70991-8

Privacy-Preserving Data Mining

Models and Algorithms

Edited by

Charu C. Aggarwal
IBM T.J. Watson Research Center, USA

and

Philip S. Yu
University of Illinois at Chicago, USA

Springer

Editors:

Charu C. Aggarwal
IBM Thomas J. Watson Research Center
19 Skyline Drive
Hawthorne NY 10532
charu@us.ibm.com

Philip S. Yu
Department of Computer Science
University of Illinois at Chicago
854 South Morgan Street
Chicago, IL 60607-7053
psyu@cs.uic.edu

Series Editors
Ahmed K. Elmagarmid
Purdue University
West Lafayette, IN 47907

Amit P. Sheth
Wright State University
Dayton, Ohio 45435

ISBN 978-0-387-70991-8 e-ISBN 978-0-387-70992-5
DOI 10.1007/978-0-387-70992-5

Library of Congress Control Number: 2007943463

© 2008 Springer Science+Business Media, LLC.
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

In recent years, advances in hardware technology have lead to an increase in the capability to store and record personal data about consumers and individuals. This has lead to concerns that the personal data may be misused for a variety of purposes. In order to alleviate these concerns, a number of techniques have recently been proposed in order to perform the data mining tasks in a privacy-preserving way. These techniques for performing privacy-preserving data mining are drawn from a wide array of related topics such as data mining, cryptography and information hiding. The material in this book is designed to be drawn from the different topics so as to provide a good overview of the important topics in the field.

While a large number of research papers are now available in this field, many of the topics have been studied by different communities with different styles. At this stage, it becomes important to organize the topics in such a way that the relative importance of different research areas is recognized. Furthermore, the field of privacy-preserving data mining has been explored independently by the cryptography, database and statistical disclosure control communities. In some cases, the parallel lines of work are quite similar, but the communities are not sufficiently integrated for the provision of a broader perspective. This book will contain chapters from researchers of all three communities and will therefore try to provide a balanced perspective of the work done in this field.

This book will be structured as an edited book from prominent researchers in the field. Each chapter will contain a survey which contains the key research content on the topic, and the future directions of research in the field. Emphasis will be placed on making each chapter self-sufficient. While the chapters will be written by different researchers, the topics and content is organized in such a way so as to present the most important models, algorithms, and applications in the privacy field in a structured and concise way. In addition, attention is paid in drawing chapters from researchers working in different areas in order to provide different points of view. Given the lack of structurally organized information on the topic of privacy, the book will provide insights which are not easily accessible otherwise. A few chapters in the book are not surveys, since the corresponding topics fall in the emerging category, and enough material is

not available to create a survey. In such cases, the individual results have been included to give a flavor of the emerging research in the field. It is expected that the book will be a great help to researchers and graduate students interested in the topic. While the privacy field clearly falls in the emerging category because of its recency, it is now beginning to reach a maturation and popularity point, where the development of an overview book on the topic becomes both possible and necessary. It is hoped that this book will provide a reference to students, researchers and practitioners in both introducing the topic of privacy-preserving data mining and understanding the practical and algorithmic aspects of the area.

Contents

List of Figures

List of Tables