Studies in Fuzziness and Soft Computing 304

For further volumes:
http://www.springer.com/series/2941

Zongmin Ma and Li Yan (Eds.)

Advances in Probabilistic Databases for Uncertain Information Management

Springer

Editors
Zongmin Ma
College of Information Science and Engineering
Northeastern University
Shenyang
China

Li Yan
School of Software
Northeastern University
Shenyang
China

ISSN 1434-9922 ISSN 1860-0808 (electronic)
ISBN 978-3-642-37508-8 ISBN 978-3-642-37509-5 (eBook)
DOI 10.1007/978-3-642-37509-5
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013935339

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Databases are designed to support the data storage, processing, and retrieval activities related to data management in information systems. Database management systems provide efficient task support and tremendous gain in productivity is hereby accomplished using these technologies. In addition, being the de-facto standard for data representation and exchange over the Web, XML (Extensible Markup Language) has been widely and deeply applied in many business, service, and multimedia applications, and a large volume of data is managed today directly in XML format.

While traditional database models provide powerful data modeling and data processing capabilities, they may suffer from some inadequacy of necessary semantics. One of the major areas of database research has been the continuous effort to enrich existing database models with a more extensive collection of semantic concepts in order to satisfy the requirements in the real-world applications. In real-world applications, information is often imperfect, and human knowledge and natural language have a big deal of imprecision and vagueness. Traditional database models assume that the models are a correct reflection of the world and further assume that the stored data is known, accurate and complete. It is rarely the case in real life that all or most of these assumptions are met. One of some inadequacy of necessary semantics that traditional database models often suffer from can hereby be generalized as the inability to handle imprecise and uncertain information. For this reason, imprecise and uncertain data have been introduced into databases for imperfect information processing by applying fuzzy logic, probability, and more generally soft computing.

It is crucial for databases to explicitly represent and process imprecise and uncertain data. This is because databases have been extensively applied in many application domains which may have a big deal of imprecision and vagueness. Imprecise and uncertain data can be found, for example, in the integration of data sources and data generation with nontraditional means (e.g., automatic information extraction and data acquirement by sensor and RFID).

Probabilistic theory can bridge the gap between human-understandable soft logic and machine-readable hard logic, and has been a crucial means of implementing

intelligent data processing and intelligent information systems. In order to deal with probabilistic information in databases, currently the research and development of probabilistic data management are attracting an increased attention.

This book covers a fast-growing topic in great depth and focuses on the technologies and applications of probabilistic data management. It aims to provide a single account of current studies in probabilistic data management. The objective of the book is to provide the state of the art information to researchers, practitioners, and graduate students of information technology, and at the same time serving the information technology professional faced with non-traditional applications that make the application of conventional approaches difficult or impossible.

This book consists of six chapters. The first two chapters focus on probabilistic data management in the context of databases, discussing probabilistic spatiotemporal databases and probabilistic object-oriented databases with fuzzy measures, respectively. The next two chapters present probabilistic data management in the context of XML. The final two chapters focus on probabilistic data management in other frameworks, covering uncertain and imprecise multidimensional data streams in OLAP and tractable probabilistic description logic programs for the Semantic Web.

Chapter 1 focuses on research in probabilistic spatiotemporal databases and presents an overview on this topic. Particularly, for the results about probabilistic spatiotemporal databases using the SPOT (Spatial PrObabilistic Temporal) approach, this chapter provides a uniform overview. Also the chapter presents numerous interesting research problems using the SPOT framework for probabilistic spatiotemporal databases that await further work.

Chapter 2 concentrates on modeling probabilistic evens with fuzzy measures in the object-oriented database model. Instead of crisp probability measures or interval probability measures of objects and classes, fuzzy sets are applied to represent imprecise probability measures. A probabilistic object-oriented database model with fuzzy measures is introduced, which incorporates fuzzy probability degrees to handle imprecision and uncertainty. Based on the proposed probabilistic object-oriented database model, several major semantic relationships of objects and classes, including equivalent object relationships, object-class relationships and subclass-superclass relationships, are investigated.

Chapter 3 aims to model and manage various kinds of uncertain data in probabilistic XML. The chapter reviews the literature on probabilistic XML. Specifically, this chapter discusses the probabilistic XML models that have been proposed and the complexity of query evaluation therein. Also the chapter discusses other data-management tasks for probabilistic XML like updates and compression, as well as systemic and implementation aspects.

Chapter 4 surveys a few applications in sensor networks, ubiquitous computing, and scientific databases that require managing uncertain and probabilistic data. The chapter presents two approaches to meeting this requirement. The first approach is proposed for a rich treatment of probability distributions in the system, in particular the SPO framework and the SP-algebra. The second approach stays closer to a

traditional DBMS, extended with tuple probabilities or attribute probability distributions, and studies the semantics and efficient processing of queries.

Chapter 5 introduces a novel approach for tackling the problem of OLAPing uncertain and imprecise multidimensional data streams via novel theoretical tools that exploit probability, possible-worlds and probabilistic-estimators theories. The result constitutes a fundamental study for this scientific field that, behind to elegant theories, is relevant for a plethora of modern data stream applications and systems that are more and more characterized by the presence of uncertainty and imprecision.

Chapter 6 proposes tractable probabilistic description logic programs (dl-programs) for the Semantic Web, which combine tractable description logics (DLs), normal programs under the answer set and the well-founded semantics, and probabilities. The chapter first provides novel reductions of tight query processing and of deciding consistency in probabilistic dl-programs under the answer set semantics to the answer set semantics of the underlying normal dl-programs. Based on these reductions, the chapter then introduces a novel well-founded semantics for probabilistic dl-programs, called the total well-founded semantics. The chapter presents an anytime algorithm for tight query processing in probabilistic dl-programs under the total well-founded semantics. It is also shown that tight literal query processing in probabilistic dl-programs under the total well-founded semantics can be done in polynomial time in the data complexity and is complete for EXP in the combined complexity. Finally, the chapter describes an application of probabilistic dl-programs in probabilistic data integration for the Semantic Web.

Acknowledgements

We wish to thank all of the authors for their insights and excellent contributions to this book and would like to acknowledge the help of all involved in the collation and review process of the book. Thanks go to all those who provided constructive and comprehensive reviews. Thanks go to Janusz Kacprzyk, the series editor of Studies in Fuzziness and Soft Computing, and Thomas Ditzinger, the senior editor of Applied Sciences and Engineering of Springer-Verlag, for their support in the preparation of this volume.

Northeastern University, China
February 2, 2013

Zongmin Ma and Li Yan

Contents