FUZZY DATABASES
Principles and Applications

INTERNATIONAL SERIES IN INTELLIGENT TECHNOLOGIES

Prof. Dr. Dr. h.c. Hans-Jürgen Zimmermann, Editor
European Laboratory for Intelligent
Techniques Engineering
Aachen, Germany

Other books in the series:

Applied Research in Fuzzy Technology
by Anca L. Ralescu

Analysis and Evaluation of Fuzzy Systems
by Akira Ishikawa and Terry L. Wilson

Fuzzy Logic and Intelligent Systems
edited by Hua Li and Madan Gupta

Fuzzy Set Theory and Advanced Mathematical Applications
edited by Da Ruan

FUZZY DATABASES
Principles and Applications

by

Frederick E. Petry
Tulane University, USA

with contribution by

Patrick Bosc
Université de Rennes, France

KLUWER ACADEMIC PUBLISHERS
Boston / Dordrecht / London

Distributors for North America:
Kluwer Academic Publishers
101 Philip Drive
Assinippi Park
Norwell, Massachusetts 02061 USA

Distributors for all other countries:
Kluwer Academic Publishers Group
Distribution Centre
Post Office Box 322
3300 AH Dordrecht, THE NETHERLANDS

Library of Congress Cataloging-in-Publication Data

A C.I.P. Catalogue record for this book is available
from the Library of Congress.

ISBN-13: 978-1-4612-8566-3 e-ISBN-13: 978-1-4613-1319-9
DOI: 10.1007/978-1-4613-1319-9

CONTENTS

PREFACE

This volume presents the results of approximately 15 years of work from researchers around the world on the use of fuzzy set theory to represent imprecision in databases. The maturity of the research in the discipline and the recent developments in commercial/industrial fuzzy databases provided an opportunity to produce this survey. In this introduction we will describe briefly how fuzzy databases fit into the overall design of database systems and then overview the organization of the text.

FUZZY DATABASE LANDSCAPE

The last five years have been witness to a revolution in the database research community. The dominant data models have changed and the consensus on what constitutes worthwhile research is in flux. Also, at this time, it is possible to gain a perspective on what has been accomplished in the area of fuzzy databases. Therefore, now is an opportune time to take stock of the past and establish a framework. A framework should assist in evaluating future research through a better understanding of the different aspects of imprecision that a database can model [1].

It is becoming evident on the commercial side that we are well within the era of the third generation of databases. (The first was graph-oriented models and the second relational systems.) Perusal of the proceedings of present-day conferences will strengthen this observation. The trend is undeniably in the direction of object-oriented databases. Deductive databases are also important, but it appears that their contribution will be in the form of logic-enhanced object-oriented databases.

Figure 1 illustrates a general view of the overall environment in which databases are designed. Enterprises are either precise or vague. The database extension that represents the enterprise is either precise or imprecise. Query languages are designed to express the user's retrieval requests in either a crisp manner or not. It should be noted that there is no general agreement on the use of terms such as " imprecise, vague, uncertain, etc. "

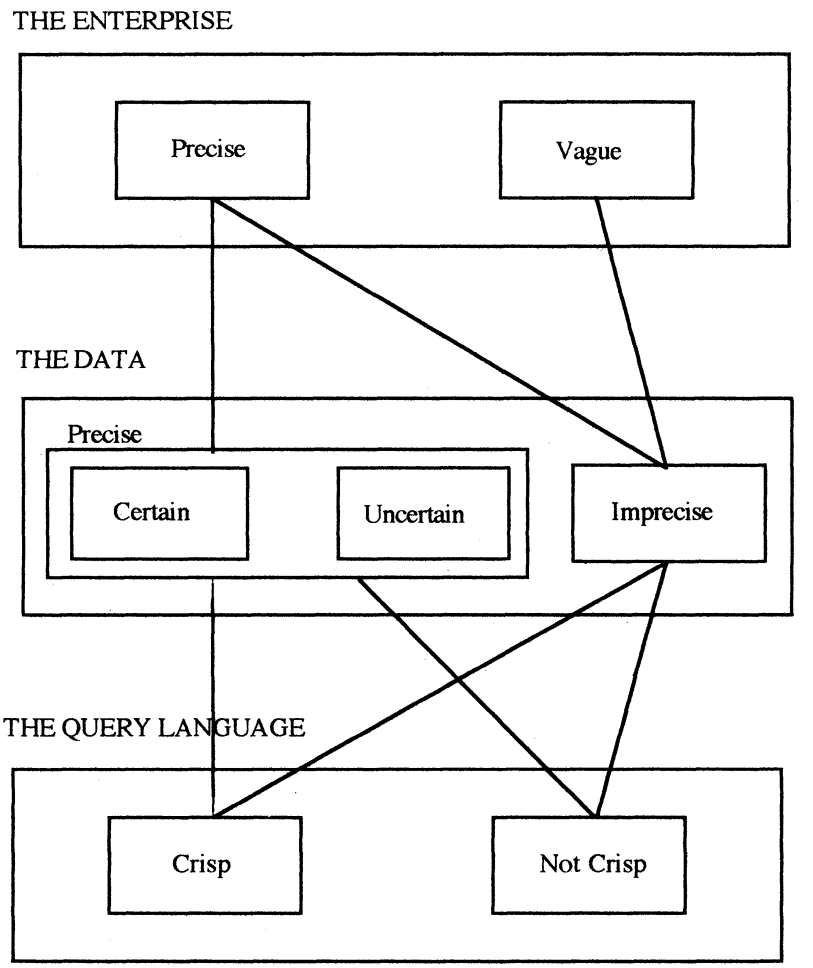

Figure 1: The Fuzzy Database Landscape

Databases are one form of modeling aspects of the real world. The specific segment of the real world which a specific database models is called the enterprise. Nearly all present databases model enterprises that are crisp. A crisp enterprise is one that is highly quantifiable - all relationships are fixed and all attributes have one value.

The case of precise enterprise-precise data includes virtually all database systems in widespread use. The potential payoff of exploiting the subcase in which the data, while precise, is considered uncertain is largely yet unexplored. If the query language is crisp or not, the issue is whether a particular data item matches a query term when it is not identical to the term. It was the precise enterprise and imprecise data that inspired one of the earliest seminal efforts on uncertainty in databases.

Certain enterprises are not crisp, but vague. For example, the family relationships among the world's languages can be fixed only in an arbitrary manner and in such enterprises, values are not necessarily crisp. Examples for which crisp values are not realistic include an international investment database containing, for each country, attributes such as "strength of judicial system", "extent of federal control over private enterprise", and "stability of government."

The lack of commercial/industrial impact of the research in fuzzy databases has been of concern [2]. It should be pointed out that other methods of uncertainty management have suffered similar fates. Most will agree that databases, like any model, are imperfect realizations of the artifact modeled. Nevertheless, there is not yet on the horizon a system having potential widespread use that incorporates a fully general means for representing or manipulating uncertainty. However a number of more specialized systems are making their appearance leading to the belief that successful commercialization of fuzzy database approaches is beginning to emerge.

ORGANIZATION OF VOLUME

This volume is organized into three major parts. The first part provides background and introductory material in the first two chapters. The next three chapters covering the major approaches in fuzzy databases comprise the second part. Finally the last chapter discusses commercial fuzzy databases.

In chapter 1 the basic concepts of databases are reviewed and a brief discussion of the major database models is provided. Then a variety of approaches to incomplete and inexact data is surveyed. These non-fuzzy representations help in understanding much of the context in which fuzzy database development has taken place. The second chapter contains an introduction to the formalisms of relational databases and their design. This is needed as the major fuzzy databases have been developed as extensions of the relational model. Finally the chapter has an overview of basic fuzzy set theory and the specific aspects commonly used in fuzzy database models.

The main approaches in fuzzy databases are covered in chapters 3, 4 and 5. Chapters 3 and 4 discuss the extensions to the major database paradigm - the relational model. Although there is a clear movement toward object-oriented models, the major developments in fuzzy databases have occurred during the past 15 years during which time relational database systems were dominant. Chapter 3

covers the use of similarity and proximity measures as the fuzzy techniques used to extend the data model; chapter 4 introduces the use of possibility theory approaches in the relational model. Both chapters discuss basic extensions to the data model, querying approaches and functional dependencies. Various other topics include implementation issues, information measures and database security. Chapter 5 surveys alternative fuzzy data models and other approaches. These include fuzzy extensions to the entity-relationship model, the IFO model and the network data models. A number of object-oriented extensions are also discussed. The use of fuzzy data modeling in geographical information systems (GIS) is described. This is an area in which there is a significant perceived need for representing imprecision at several levels. Finally the use of rough sets to represent imprecision is described and rough and fuzzy rough relational data models are presented.

The final chapter examines the issues related to commercialization of fuzzy databases. In the context of the issues relevant to practical use of fuzzy databases, several specific industrial/commercial products and applications are described. These include approaches to developing fuzzy front-end systems and special-purpose systems incorporating fuzziness.

REFERENCES

1. B. Buckles and F. Petry. "Fuzzy Databases in the New Era" *Proceedings of FUZZ-IEEE/IFES'95 Workshop on Fuzzy Database Systems and Information Retrieval*, 85-91, 1995.

2. F. Petry,"Information Systems for the 21st Century: A Fuzzy-Oriented Manifesto," Invited Presentation, *Proceedings IFSA 95*, 1995.

ACKNOWLEDGEMENTS

This volume must recognize the influence of Lotfi A. Zadeh. Of course Lotfi indeed established the entire field of fuzzy sets, but furthermore he has had long-standing interest and insight into the applications of fuzzy set theory to databases.

First and foremost I would like to acknowledge my collaborator in this effort- Patrick Bosc. Chapter 4 on possibility-based databases is Patrick's work as he is a foremost researcher in this area. Additionally we wish to thank Dr. Hans Zimmermann for inviting us to produce this book. We wish to acknowledge our debt to all of the researchers in the area of fuzzy databases. Their influence on this volume is profound, based on both their publications and the many discussions in which we have engaged at meetings around the world. We have attempted to be comprehensive but not necessarily encyclopedic in describing their efforts and hope we have not unintentionally omitted anyone.

We would like to thank colleagues Drs. Theresa Beaubouef and Roy George who have reviewed various chapters. Also we wish to acknowledge Dr. Miguel Zuniga, Linda Tolbert and Alice Leblanc for their assistance in preparation of various parts of the text.

Finally we wish to thank our families for their patience when we needed to devote spare time in development of this volume.

Frederick E. Petry
Center for Intelligent and Knowledge-Based Systems (CIAKS)
School of Engineering
Tulane University, USA

Patrick Bosc
IRISA, École Nationale Supérieure de Sciences Appliquées
et de Technologie (ENSSAT)
Université de Rennes, France

FUZZY DATABASES
Principles and Applications