

Preface

1 Introduction

Nowadays, data management constitutes a blossoming domain that expands rapidly: the development of numeric applications and mobile devices generates a profusion of data in numerous activity areas. Data sources multiply at a sustained pace: social networks, sensor networks, user data, business data, web data. Data sets are getting huge, more and more open, often pervasive, and their exploitation becomes a major societal challenge.

In this context, research issues are numerous and diverse, in particular those concerning access to information: customization, preference queries, data summaries, database mining, heterogeneous and complex data handling, data integration, uncertain data management.

A crucial issue in database research these days is how to make systems more *flexible* and more *human-centric*. This implies, among other aspects, taking into account the preferences of the users as well as their context, to be able to deal with uncertain data, to exploit metadata such as ontologies in order to answer queries, to devise systems that exhibit a cooperative behavior, etc. Notice that similar problems arise in information retrieval as well. Besides classical database tools that need to be extended, knowledge discovery techniques can also be leveraged to make the content (and the structure) of a database more intelligible to users.

The application of fuzzy set theory to the database domain is an already old story that started in the late 1970s. Indeed, it appeared very early that fuzzy sets constitute an intuitive and powerful tool to model and handle gradual concepts in the context of databases (e.g., vague predicates involved in flexible queries, approximate functional dependencies, imprecise values, user profiles). Pioneering works are those by V. Tahani, H. Prade, P. Bosc, B. Buckles and F. Petry, M. Zemankova and A. Kandel, M. Umamo, J. Kacprzyk, to cite a few.

Patrick Bosc played a leading role in this community from the very beginning. His position is confirmed with many publications, plenary Lectures at many prestigious conferences, but also with the response we have received while sending invitations to contribute to this volume dedicated to Patrick, to celebrate his retirement. Patrick, an expert in database theory, quickly recognized a potential of fuzzy set theory to make standard database concepts and tools closer to human way

of thinking. One of his great achievements was the development of the SQLf language, which combined a highly popular query language with elements of fuzzy logic in a very comprehensive way, covering all the details of SQL.

The topics listed above (preference queries, uncertainty management, cooperative answering, etc.) are now recognized in mainstream database research, but there is still some effort to make in order to convince the researchers from that community that fuzzy sets are an appropriate tool for tackling these issues. This volume is hopefully a step toward that goal.

2 Structure of the Volume

2.1 *Bipolar Preference Queries*

Bipolarity refers to the propensity of the human mind to reason and make decisions on the basis of positive and negative affects. Positive information states what is possible, satisfactory, permitted, desired, or considered as being acceptable. On the other hand, negative statements express what is impossible, rejected, or forbidden.

The first part of the volume, devoted to bipolar preference queries, includes four chapters. The first, by D. Dubois and H. Prade, revisits the modeling of the connective *if possible* in requirements of the form “*A and if possible B*.” The authors mainly distinguish between two types of understanding: either (i) *A* and *B* are requirements of the same nature and are viewed as constraints with different levels of priority, or (ii) they are of different nature (only *A* induces constraint(s) and *B* is only used for breaking ties among items that are equally satisfying *A*). The authors show that the two views are related to different types of bipolarity, and discuss them in relation with possibilistic logic. The disjunctive dual of the first view (“*A or at least B*”) is then presented in this logical setting.

The second chapter, by T. Matthé, J. Nielandt, S. Zadrozny, and G. De Tré, provides an overview of two approaches to bipolar preferences. Both approaches use pairs of satisfaction degrees as an underlying framework, but have different semantics, and hence lead to different operators for criteria evaluation, ranking, aggregation, etc.

In the third chapter, by L. Liétard, D. Rocacher, and N. Tamani, the authors also consider two different types of bipolar preferences: conjunctive ones (*if possible*) and disjunctive ones (*or else*) and show that both of them can be interpreted in a hierarchical way. They then introduce a general form of fuzzy bipolar conditions and extend the operators of relational algebra so as to give them a bipolar semantics and make them work on bipolar fuzzy relations.

In the fourth chapter, J. Kacprzyk and S. Zadrozny investigate different aspects and interpretations of bipolarity in database querying. The authors analyze various ways to deal with bipolarity, recast them in a unified perspective, and clarify them with respect to conceptual, algorithmic, and implementation related aspects.

See also the chapter by Buche et al. in the third part of the volume (cf. [Sect. 2.3](#)) that deals with bipolar queries to uncertain databases using ontologies.

2.2 Ontology-Based Data Access

The second part of the volume, made of two chapters, is devoted to Ontology Based Data Access (OBDA), whose aim is to use an ontology to mediate the access to data sources. The added value of OBDA, w.r.t. accessing a data source directly, is, on the one hand, that the ontology provides a semantic account of the information stored in the data source. On the other hand, the answer to user queries may be enriched by exploiting the constraints expressed by the ontology, thus overcoming incompleteness that may be present in the data.

The first chapter, by U. Straccia, discusses the problem of evaluating ranked top-k queries in the context of ontology mediated access over relational databases. An ontology layer is used to define the relevant abstract concepts and relations of the application domain, while facts with associated score are stored into a relational database. Queries are conjunctive queries with ranking aggregates and scoring functions. The results of a query may be ranked according to the score and the problem is to find efficiently the top-k ranked query answers.

The second chapter, by J. R. Campana, J. M. Medina, and M. A. Vila, presents a schema and a transformation algorithm to store OWL ontologies in Object Relational Databases. The database schema makes it possible to represent an ontology structure, while the transformation algorithm creates an appropriate schema to store its instances preserving all information. The approach presented enables using instance data of imprecise nature, mostly fuzzy numerical data.

2.3 Uncertain Databases

In the late 1970s, Database Researchers (notably E. F. Codd and W. Lipski) started investigating the issue of extending the relational database model so as to represent unknown (null) values. Since then, many authors have proposed diverse approaches to the modeling and handling of databases involving uncertain or incomplete data. In particular, the last two decades have witnessed an explosion of research on this topic, with many chapters devoted to probabilistic databases (and also a few dealing with possibilistic databases, notably by Bosc et al.).

This third part of the volume, devoted to uncertain databases, includes three chapters. The first, by T. Beaubouef and F. Petry, discusses rough set, fuzzy rough set, and intuitionistic rough set approaches and how to incorporate uncertainty management using them in the relational database model.

The second chapter, by P. Buche, S. Destercke, O. Haemmerlé, and R. Thomopoulos, proposes an approach to query a database where the user

preferences can be bipolar (cf. [Sect. 2.1](#) above) and the data stored in the database can be uncertain. Query results are then completely ordered with respect to these bipolar preferences, giving priority to constraints over wishes. Furthermore, the authors consider user preferences expressed on a domain of values which is not “flat,” but contains values that are more specific than others according to the *kind of relation*.

The third chapter, by J. Pons, C. Billiet, O. Pons and G. De Tré, deals with imperfect data in temporal databases. The authors first provide an overview of the basic concepts and issues related to the modeling of time as such or in relational database models and the imperfections that may arise during or as a result of this modeling. Then, they present a novel technique for handling some of these imperfections.

2.4 Flexible Queries over Nonstandard Data

The notion of fuzziness in information retrieval has been long recognized. Indeed, broadly meant fuzziness may be existent in representing documents, in representing queries, in matching the representations of documents to the representations of queries, in evaluating the retrieved documents by users, and even in evaluation of the performance of the retrieval system. The first chapter of this fourth part of the volume, by G. Pasi, G. Bordogna, and G. Psaila, presents a unifying model of flexible queries with distinct semantics of search terms weights. When querying documents archives, there is often the need to specify importance weights of the search terms that define flexible selection conditions on documents representation. Several interpretations of the semantics of these weights have been proposed within distinct information retrieval models. In this contribution, the authors define a unifying model of information retrieval based on a vector p -norm, where importance weights with distinct semantics can be specified in flexible queries.

Social networks have rapidly become an important technology in our digital-based information intense world. In addition to enabling people from all over the world to connect with each other, they provide vast sources of information about the individual participants in the network. Each of these participants can view a kind of database containing information about themselves. Thus, a social network can be viewed as a kind of network of databases. The second chapter of this fourth part of the volume, by R. Yager, discusses fuzzy relationships and their role in modeling weighted social relational networks. The author describes how the idea of “computing with words” can provide a bridge between a network analyst’s linguistic description of social network concepts and the formal model of the network. The idea of vector-valued nodes is introduced and then the basic elements of a technology of social network database theory are presented.

2.5 *Fuzzy Knowledge Discovery and Exploitation*

The paradigm of cooperative answering originated from the works concerning natural-language question-answering in the early 1980s. One of the aims of such works is to prevent systems from producing “there is no result” when a query fails. In the first chapter of this fifth part, G. Smits, O. Pivert, and A. Hadjali propose a unified framework for cooperative answering, that relies on a fuzzy-cardinality-based summary of the database. The authors show how this type of summary can be efficiently used to explain failing queries or to revise queries returning a plethora answer set.

The last chapter, by M. Q. Flores, F. Del Razo, A. Laurent, and N. Sicard deals with the parallelization of fuzzy database mining algorithms. The authors discuss why a parallel approach is crucial to tackle the problem of scalability and optimal performance in the context of database mining. They then present parallel algorithms on multi-core architectures devoted to four knowledge discovery paradigms, namely fuzzy association rules, fuzzy clustering, fuzzy gradual dependencies, and fuzzy tree mining.

The editors wish to thank the following reviewers for their invaluable help: Troels Andreasen, Isabelle Bloch, Gloria Bordogna, Patrice Buche, Jesus Roque Campana, Davide Ciucci, Sébastien Destercke, Guy De Tré, Didier Dubois, Michel Grabisch,ALLEL Hadjali, Olgierd Hryniewicz, Hélène Jaudoin, Janusz Kacprzyk, Donald Kraft, Ludovic Liétard, Zongmin Ma, Christophe Marsala, Arnaud Martin, Adam Niewiadomski, Henri Prade, Grégory Smits, Umberto Straccia, Eulalia Szmidt, Laurent Ughetto, Peter Vojtas.

Olivier Pivert
Sławomir Zadrozny

Flexible Approaches in Data, Information and
Knowledge Management

Pivert, O.; Zadrozny, S. (Eds.)

2014, XII, 320 p. 56 illus., Hardcover

ISBN: 978-3-319-00953-7