

Fuzzy Databases — Principles and Applications

by Frederick E. PETRY

(with chapter contribution by Patrick BOSCH)

Kluwer Academic Publishers, Dordrecht, 1996, 226 pages

If we except the early monograph by Zemankova and Kandel (1984) and the recent collection of articles edited by Bosch and Kacprzyk (1995), this book is one of the very first books solely devoted to fuzzy databases, and the first treatise reporting about the researches in this area which have been developed for about 20 years now.

If we take apart borderline topics such as multiple sources data fusion or data mining, fuzzy databases literature has been concentrating mainly on three issues, namely i) the processing of flexible queries; ii) the handling of imprecise, uncertain or fuzzy data; iii) the study of fuzzy (functional) dependencies.

The present volume of the International Series in Intelligent Technologies provides a well-informed account of these different lines of research in a limited number of pages (226 pages). Each chapter is followed by a very rich bibliography of papers cited in the text. It can be divided in three main parts.

The first part, made of two chapters, introduces to database fundamentals (principles, design, models, null values and incomplete information). It gives the necessary background on, 1) the relational database framework in which most of the present fuzzy approaches take place (basic concepts, functional dependencies and normal forms, query languages are surveyed), and 2) on fuzzy set theory (set operations, similarity and proximity relations, possibility theory).

The second part is made of three chapters which cover the main approaches: i) similarity-based models, ii) possibility-based models, and iii) other approaches. Similarity-based approaches were initiated by Buckles and the author of this monograph about fifteen years ago and have been developed further from this line. The idea is to attach a fuzzy similarity (i.e., reflexive, symmetric and max-* transitive for some operation *) relation to each attribute domain and to allow for set-valued attribute values (with the idea that the values in the set are somewhat similar). Then the classical notion of redundant tuples (based on perfect identity) can be weakened using the idea of approximate similarity. It is also possible to replace similarity relations by proximity relations (which have no transitivity properties). From the querying point of view this framework enables the system, for instance, to retrieve items which agree to some extent with a definite set of items with respect to some attributes. An extension of the relational calculus for handling similarity is presented, which thus copes with a form of uncertainty about the precise meaning of attribute values (by taking also into account similar attribute values in the evaluation). Fuzzy functional dependencies, in relation with design concerns, are also introduced in this framework. Lastly, the chapter reports on the application to the protection of security of data (roughly speaking, the idea is to control the access to protected values as well as to the values which are sufficiently similar).

The chapter on possibility-based models is a contribution due to Patrick Bosc, a specialist of this type of approach. This chapter with a large coverage, starts with the representation of imprecise and uncertain data (including null values) by means of possibility distributions, and the evaluation of flexible queries addressed to a database including such pieces of data. The evaluation machinery is based on a fuzzy pattern matching technique yielding to what extent it is possible on the one hand, and to what extent it is certain on the other hand, that a tuple satisfies a given query. Then the chapter focuses on the particular case of flexible queries directed to precise data (in this case the possibility and certainty degrees computed in the general case coincide). Special attention is dedicated to the evaluation of quantified statements (e.g., "most X's are A's"), and to possible definitions of the division operation in the fuzzy set framework. A fuzzy extension of the SQL language is presented. Implementation aspects and query processing are then discussed. The end of the chapter also contains an overview of works on fuzzy functional dependencies (usually of the type: "the more similar two tuples with respect to attributes in X are, the more similar they should be for attributes in Y"), discussing their possible use in databases containing precise information.

The chapter on alternative database models give brief accounts of works proposing fuzzy set-based extensions of various database approaches: entity relationship models, object-oriented models (using fuzzy similarity relations, or other fuzzy set and possibility theory tools), rough set-based approach (based on the idea of indiscernibility relation between objects with respect to a given set of attributes), and geographical information systems.

The third part of the book, which is made one chapter, is devoted to commercial applications. The presence of this chapter at the end of the book is clearly a valuable addition and may be also viewed as announcing the blossoming of more applications in the future. All the applications reported mainly focus on the processing of fuzzy (i.e., flexible) queries. However it seems that for the moment very little effort is devoted to the elicitation of the meaning intended by the user when he/she addresses a fuzzy query to a database system. It is usually for expressing preferences among attributes values, and priorities between elementary requirements (vague queries are also a way for avoiding empty sets of answers by expressing possible relaxations of the requirements). Moreover it may be also important to understand what kind of conjunctions (logical or compensatory) the user has in mind in case of compound queries. These problems call for more works on the interface with the user, which may be importance for the development of more applications.

On the whole the book provides a well-documented and comprehensive view of the on-going research in fuzzy databases.

Henri PRADE

References

- Zemankova-Leech M., Kandel A. "Fuzzy Relational Data Bases — A Key to Expert Systems", Verlag TÜV Rheinland, Köln, 1984
- Bosc P., Kacprzyk J. (Eds.) "Fuzziness in Database Management Systems", Physica-Verlag, Heidelberg, 1995