Chapter 3.16
Rough Sets:
A Versatile Theory for Approaches to Uncertainty Management in Databases

Theresa Beaubouef
Southeastern Louisiana University, USA

Frederick E. Petry
Naval Research Laboratory, USA

ABSTRACT

This chapter discusses ways in which rough-set theory can enhance databases by allowing for the management of uncertainty. Rough sets can be integrated into an underlying database model, relational or object oriented, and also used in the design and uerying of databases, because rough-sets are a versatile theory, theories. The authors discuss the rough relational databases model, the rough object-oriented database model, and fuzzy set and intuitionistic set extensions to each of these models. Comparisons and benefits of the various approaches are discussed, illustrating the usefulness and versatility of rough sets for uncertainty management in databases.

INTRODUCTION

Rough-set theory has become well established since first introduced by Pawlak in the 1970s. It is based on two simple concepts: indiscernibility and approximation regions. Rough-set theory is a formal theory, mathematically sound. It has been applied to several areas of research such as logic and knowledge discovery, and has been implemented in various applications in the real world. Because of rough sets' ability to define uncertain things in terms of certain, definable things, it is a natural mechanism for integrating real-world uncertainty in computerized databases. Moreover, other uncertainty-management techniques may be combined with rough sets to offer even greater uncertainty management in databases. This chapter discusses how rough-sets theory can be applied to several areas of databases including design, modeling, and querying. Both relational and object-oriented databases benefit from rough-set techniques, and when combined with fuzzy or intuitionistic sets, these databases are very rich in the modeling of uncertainty for real-world enterprises.

BACKGROUND

Rough-set theory, (Pawlak, 1984, 1991) is a technique for dealing with uncertainty. The following concepts are necessary for rough sets:

- U is the *universe*, which cannot be empty,
- R is the *indiscernibility relation*, or equivalence relation.
- $A = (U,R)$, an ordered pair, is called an *approximation space*.
- $[x]_R$ denotes the equivalence class of R containing x, for any element x of U.
- *Elementary sets* in A—the equivalence classes of R.
- *Definable set* in A—any finite union of elementary sets in A.

A given approximation space defined on some universe U has an equivalence relation R imposed upon it, partitioning U into equivalence classes called elementary sets that may be used to define other sets in A. Given that $X \subseteq U$, X can be defined in terms of the definable sets in A by:

lower approximation of X in A is the set
$\underline{R}X = \{x \in U \mid [x]_R \subseteq X\}$

upper approximation of X in A is the set
$\overline{R}\,X = \{x \in U \mid [x]_R \cap X \neq \varnothing\}$.

The major rough-set concepts of interest are the use of an indiscernibility relation to partition domains into equivalence classes and the concept of lower and upper approximation regions to allow the distinction between certain and possible, or partial, inclusion in a rough set. Indiscernibility is the inability to distinguish between two or more values. For example, the average person describing the color of a suspect's hair may say that it is "brown," when it actually is dark brown. As it turns out, "brown" is probably good enough for helping the police identify the suspect. However, a beautician who specializes in hair color will find it important to discern between the various shades of brown. Indiscernibility can also arise from lack of precision in measurement, limitations of computational representation, or the granularity or resolution of the sampling or observations

ROUGH SETS IN DATABASE DESIGN

Beaubouef and Petry (2004a, 2005a) introduce a rough-set design methodology for databases. Conceptual modeling is accomplished through rough entity-relationship modeling and then for relational databases, rough normalization is discussed. The process of normalization makes use of the concept of rough functional dependencies.

We must first design a database using some type of semantic model. We use a variation of the entity-relationship diagram that we call a fuzzy-rough E-R diagram. This diagram is similar to the standard E-R diagram in that entity types are depicted in rectangles, relationships with diamonds, and attributes with ovals. However, in the fuzzy-rough model, it is understood that membership values exist for all instances of entity types and relationships. Attributes that allow values where we want to be able to define equivalences are denoted with an asterisk (*) above the oval. These values are defined in the indiscernibility relation, which is not actually part of the database design, but inherent in the fuzzy-rough model.