Book Reviews

Stéphane P. Demri and Ewa S. Orłowska, **Incomplete Information: Structure, Inference, Complexity**, Springer-Verlag, Berlin Heidelberg 2002. US$ 101.00, pp. 450, ISBN 3540419047, hardcover. Dimensions (in inches): $9.5 \times 6.3 \times 1.1$.

Zdzisław Pawlak introduced his model for information systems in 1982, tough he had presented many essential ideas concerning information systems already in the early seventies. An information system in the sense of Pawlak consists of a set of objects, called the universe, and a set of attributes A. Further, for each attribute $a \in A$, a set V_a consisting of the values of the attribute a is attached. Every attribute $a \in A$ can be viewed as a mapping $U \to V_a$ and the image $a(x)$ is the value of the attribute a for the object x. Pawlak's model thus differs slightly from the better known relational model of databases by Edgar Codd in which rows, called tuples, are considered as elements of the product of value sets of attributes. More precisely, a relation r of a relation schema $R(A)$, where R is the name of the relation and A is the set of attributes, is a set of tuples t and $t[a]$ denotes the value of the attribute a for the tuple t.

While Codd's model is widely used in contemporary database design, information systems in the sense of Pawlak have became popular in rough-set-based data analysis and data mining. The fundamental idea in Pawlak's information systems is that each subset $B \subseteq A$ of attributes determines a so-called indiscernibility relation $\mathrm{ind}(B)$ which is defined so that two objects x and y of the universe U are B-indiscernible if their values for all attributes in the set B are equal. With respect to indiscernibility relations dependency relations resembling functional dependencies in relational databases and implications in formal concept analysis may be defined. While in database theory dependencies describe actual connections between attributes, for example, the social security number determines the name and the address of a person, in Pawlak's information systems we may find new dependencies between attribute sets, and such dependencies may reveal new and interesting correlations. Formally, a set $C \subseteq A$ is said to be dependent on a set $B \subseteq A$, denoted $B \to C$, if $\mathrm{ind}(B) \subseteq \mathrm{ind}(C)$. This means that if $B \to C$ holds, then whenever two objects are B-indiscernible, they must be also C-indiscernible. Thus, we may write if-then rules such that if an object has certain values for attributes in B, its values for the C-attributes are completely determined.

Pawlak has also introduced the notion of rough set approximations which naturally arise from information systems. The idea is that in many situations our knowledge of objects is incomplete, and it can be viewed as being restricted by some indiscernibility relations. Indiscernibility relations are equivalences interpreted so that two objects are indiscernible if they cannot be distinguished by their properties which are known to us. Because the correspondence between equivalences and partitions is one-to-one, every indiscernibility relation induces a partition of the

Studia Logica (2006)
DOI: 10.1007/s11225-006-9018-5

universe consisting of blocks of indistinguishable objects. This can be interpreted so that our ability to see objects is 'blurred' – we cannot perceive individual objects, only indiscernibility classes of them. By the induced partition we may divide objects of the universe U into three classes for every concept $X \subseteq U$: the objects that certainly are in X, the objects which possibly are in X, and the objects which certainly are not in X. More formally, the upper approximation of X consists of the equivalence classes intersecting with X, the lower approximation consists of all the blocks included in X, and the boundary of X consists of the set-theoretical difference of the upper and the lower approximation. Now, the lower approximation can be viewed as the set of elements which certainly are in X, the boundary is the actual area of uncertainty, and the elements belonging to the complement of the upper approximation are elements that certainly are not in X in view of the knowledge represented by the given indiscernibility relation. It should be also noted that approximations and boundaries are definable in the sense that they are unions of some equivalence classes. Therefore, if an indiscernibility relation is determined by some subset $B \subseteq A$ of an information system, approximations and boundaries can be described by using the values of B-attributes.

Rough set approximations determine also a 'second-order' indiscernibility relation on the set of all subsets of the universe: two sets of objects X and Y are roughly equal if their lower and upper approximations of X and Y coincide; this means that precisely the same elements possibly belong and certainly belong to the mutually roughly equal sets X and Y. The equivalence classes of this relation are called rough sets, and clearly sets belonging to the same rough set look exactly the same when viewed through the indiscernibility relation. The set of rough sets has an interesting lattice-theoretical structure. It forms a complete Stone lattice isomorphic to $\mathbf{2}^I \times \mathbf{3}^J$, where $\mathbf{n}$ is the n-element chain, I is the set of one-element indiscernibility classes, and J is the set of classes having at least two elements.

Ewa Orłowska and Pawlak generalized the notion of information systems in 1984. They introduced so-called nondeterministic information systems in which every attribute in A assigns a subset of values of V_a to objects in U, that is, each attribute is a mapping $U \to \wp(V_a)$. The systems of this kind enable us to represent many-valued attributes. For example, if a is the attribute 'knowledge of languages' and a person x speaks French and Polish, then we have $a(x) = \{\text{French}, \text{Polish}\}$. We may also represent situations in which we do not know the exact value for the attribute, but we can give a set such that the correct value is supposed to be in that set. For example, if a is the attribute 'age' and we know that a person denoted by x is in her late thirties, then we may write $a(x) = \{36, 37, 38, 39\}$. In nondeterministic information systems it is possible to define several so-called information relations reflecting distinguishability and indistinguishability of objects with respect to different attribute sets. For example, if a is the attribute 'knowledge of languages', then two objects x and y are a-similar if they have a common language, that is, $a(x) \cap a(y) \neq \emptyset$. Especially, Orłowska has defined and studied properties of a large variety of information relations of different kinds. By means of information relations, information operations of possibility and necessity generalizing rough approximations operations may be defined. Further, also operations representing impossibility and sufficiency can be defined by means of possibility and necessity operators. Depending on the properties of the underlying information relations, these operators have different kinds of properties. Also a large amount of so-called information logics based on information relations with Kripke-style semantics can be found in the literature.

The present monograph gives a systematic presentation of the theory related to the above-described relation-based knowledge representation and the underlying logics and algebras. The problem of informational representability is central to the book. The adequacy of the presented logical and algebraic systems is in most cases confirmed by the representation theorems saying that every model of a logic or every algebra of a class is properly associated with a similar structure derived from an information system. The book starts with an 11-page chapter recalling some essential mathematical prerequisites, including basic notions concerning sets, relations, and mappings. It also covers some universal algebra and lattice theory including homomorphism theorems, basic results about filters and complementation in lattices.

After the preliminary chapter, the monograph is divided into five parts. The first part is most fundamental since it recalls the essential concepts related to relation-based data representation. The authors begin by defining information systems and also two special types of them: total systems in which the values of attributes are always non-empty, and deterministic systems in which attributes may have at most one value. Some examples of various types of information systems are also given. The authors also recall the notion of property systems by Dimiter Vakarelov, consequence systems of Dana Scott, and the formal concepts by Rudolf Wille, and show that each of these can be regarded as a certain kind of an information system. A special type of information systems, called decision tables, are also shortly considered. In decision tables, there exists a distinguished attribute which is referred to the decision attribute, and the idea is that each table can be regarded as an if-then decision rule: the values of the condition attribute determine which values the decision attribute may have. L-sets were introduced by Joseph Goguen in 1967 to generalize the notion of 'traditional' fuzzy sets in which the membership function of a set is to the $[0, 1]$-interval. In L-sets the idea is that L can be an ordered set describing, for example, linguistic hedges like "very" or "quite". In the book, fuzzy information systems are such that the values of attributes are L-fuzzy subsets, where L is a double residuated lattice, that is, a lattice with additional operators of product, sum, and the corresponding residua.

The first part also contains the definitions of information relations, such as indiscernibility, similarity, and inclusion – which represent indistinguishability of objects in terms of their attributes – and diversity, orthogonality, and negative similarity representing distinguishability. For each relation type both strong and weak versions are given: 'strongness' means that objects are in a specified relation with respect to all attributes in sets, as 'weakness' means that they are related with respect to at least one attribute of the set. The essential properties of the relations are given together with some lattice-theoretical considerations about the structure of the family of all indiscernibility relations. A short section related to indiscernibility is devoted to data dependencies and reduction.

Relational frames have an important role in the present monograph. By definition, a relational frame consists of an object set and a family of information relations indexed by the power set of attributes derived from an information system. As a generalization of information frames arising from information systems, the authors define 21 different types of abstract frames and give examples of members of such systems. They also consider fuzzy counterparts of information relations.

Part one ends with a study of information operators. First the rough approximation operators, rough inclusions and equalities are considered. Then modal-like operators of possibility, necessity, impossibility, and sufficiency with respect to an arbitrary binary relation are introduced. Several correspondence results for the operators are also given in the style of modal correspondence theory. Again, fuzzy versions of approximation operators are given.

Part two is relatively short consisting only of two chapters that provide the logical background for the rest of the book. The first chapter of this part recalls the definitions and basic facts of propositional calculus. Then, general notions of first-order predicate logic and modal logic are presented, and the basic modal logics K, T, B, S4, and S5 are considered as examples. Also the data analysis logic (DAL) and the propositional dynamic logic (PDL) are defined. In the first part, the authors defined relations indexed by families of different levels of the powerset hierarchy, and here the logics with such relative accessibility relations are defined and they are called Rare-logics. The first chapter ends with a brief description of the construction of fuzzy modal logics. The second chapter is quite technical. Various model-theoretic constructions for modal logics are given including copying, filtration, restriction, and canonical frames. Then, Hilbert-style proof systems are recalled for classical propositional and modal logics. The chapter ends by recalling notions of complexity theory. Turing machines are chosen as the model of computation. However, no Turing machine 'programming' is performed in the book. Finally, several complexity classes, including NP, PSPACE, EXPTIME, and NEXPTIME are recalled.

The third part of the book is devoted to various information logics. The part is divided into three chapters where reasoning about similarity, reasoning about indiscernibility, and reasoning about knowledge in general are considered. For reasoning about similarity, three modal logics, namely the logic for reasoning about non-deterministic information (NIL), the information logic (IL), and the logic for reasoning about relative similarity (SIM), are given. The languages and semantics of these logics are defined and for each of the logics completeness theorems are presented. For NIL and IL, Hilbert-style proof systems are defined, as for SIM, which is a Rare-logic, a Rasiowa–Sikorski-style proof system is given.

The second chapter of this part studies reasoning about indiscernibility, which means that semantic structures of logics contain relations which are counterparts of indiscernibility relations arising from informations systems. The data analysis logic with local agreement (DALLA) with its Hilbert-style deduction system is considered. DALLA is a variant of DAL considered in part two. For DALLA, also a completeness theorem is given. In the second chapter, logics of relative indiscernibility and so-called LA-logics are considered. The logic IND includes a family of abstract indiscernibility relations indexed by the power set of some finite set and LA-logics can be viewed as a generalization of DALLA assuming various classes of local agreement in forms of preorders in the semantic structures. For IND and LA-logics, Hilbert-style proof systems and completeness theorems are given. The chapter ends with a short survey on a fuzzy logic of modalities.

The last chapter of part three considers reasoning about knowledge in general. Three logics are introduced: BLKO (basic logic with a knowledge operator), S5$'$ (an extension of S5), and LKO (logic with relative knowledge operators). For all these logics Hilbert-style deduction systems are defined and their completeness is proved. The idea behind BLKO is that it includes a single knowledge connective that mimics the knowledge operator K defined by means of rough approximations such that for each subset X of the universe, $K(X)$ is the area of certainty, that is, the complement of the boundary of X. The logics S5$'$ is an extension of S5 with a set of equivalences and their set-theoretical intersections. LKO can be considered as an extension of BLKO to a Rare-logic.

Complexity issues of different information logics are considered in the monograph's fourth part. The first chapter investigates relationships between standard modal logics and Rare-logics with a unique relation type, that is, logics containing a single family of relative relations. Several classes of Rare-logics are introduced and the authors show how formulas of certain Rare-logics can be transformed into formulas of standard modal logics preserving validity. At the end of the chapter so-called reducible Rare-logics are considered and some conditions for Rare-logics to be reducible are stated. The second chapter of this part studies complexities of information logics. The authors show that the logics NIL, IL, LKO, DALLA, S5$'$, SIM, and fuzzy logics of graded modalities are decidable by applying the techniques presented in part two; the decidability of BLKO was presented already in the previous part. Also some open problems are addressed – for example, the decidability of DAL.

The third chapter lists time and space complexities of different logics. The titles of the sections describe the main results of this chapter: 'IND-satisfiability is is LIN', 'NP-complete LA-logics', 'BLKO-satisfiability is NP-complete', 'NIL-satisfiability is PSPACE-complete', 'S5$'$-satisfiability is EXPTIME-complete', and 'LKO-satisfiability is EXPTIME-complete'. The final chapter presents some other complexity bounds based on known results. For example, DALLA- and LGM-satisfiability are PSPACE-complete, SIM$'$- and DAL-satisfiability are EXPTIME-hard, and IL-satisfiability is PSPACE-hard.

Part five studies algebraic approaches to information operators. It begins with some methodological issues considering the concept of informational representability. This is done by introducing the notion of Σ-frames, where Σ is the frame signature, and the method itself is referred to as the nice set proof technique. The second chapter studies algebraic structures determined by information operators.

It is shown, for instance, that the family of all indiscernibility relations of an information system forms a complete lattice and that any family of sets closed under information operators on the powerset of the universe, such as positive, negative, lower borderline, and upper borderline operators, forms a Boolean algebra. The authors also show that the Boolean algebra of all subsets of the universe with the upper approximation operators forms a Łukasiewicz algebra. Further, the family of all rough sets, that is, the family of lower-upper approximation pairs, forms a three-valued Łukasiewicz algebra. The well-known result that family of all rough sets forms a regular double Stone algebra is recalled. For three-valued Łukasiewicz algebras and for Stone algebras, representation theorems in terms of rough set algebras are given. The second chapter also studies connections between approximation spaces and Nelson algebras, as well as algebras of rough relations and fuzzy relations. The chapter ends with a study of algebraic structures of generalized approximation operators. These are defined by means of so-called rough posets that are subsets of some poset, and they form themselves complete lattices; their elements, which may be considered 'exact', are used as approximations.

The final chapter studies algebraic counterparts of the information frames defined in part one. Different types of self-maps on a Boolean algebra are introduced, and with respect to these types, 'algebraic' modal operators, dual modal operators, sufficiency operators, and dual sufficiency operators are defined. Then, 33 different classes of algebras with operators are defined and the authors develop their theory in the spirit of Jónsson and Tarski together with some examples of the members of these classes. The book ends with a short chapter discussing relationships between information logics and information algebras. These are similar to the known correspondences between modal logics and modal algebras.

The book gives a complete and systematic survey of formal methods and theories of the past 25 years of data analysis and information logics inspired by rough set theory. The structure of this book is clear. As the name of the book states, structure, inference, and complexity issues on reasoning about incomplete information are considered in their own parts together with a part considering algebraic aspects of information operators.

The book is written to be self-contained, but unfortunately this is not always the case due to the many notions and structures considered. Especially in part one, where the fundamental concepts of the theory are introduced, more examples could be given. Another problem is that many proofs of important results are omitted and often replaced with a sentence 'the proof is by an easy verification', which is in some cases misleading and not true. There are theorems that have quite complicated proofs of several pages, and at least to me these do not appear as easy verifications. In addition to this, the reader sometimes gets the impression that the authors try to cover even too many subjects, since the book contains several definitions that are not used at all.

The references section of this book is excellent and it shows that the authors have really gone deeply into the subjects of this monograph. The end of each chapter contains notes providing historical perspective on the issues of the chapter. Further, in the bibliography, the pages in which a given entry is cited, are mentioned.

The book can be recommended to researchers and graduate students interested in rough-set-style reasoning and knowledge representation. Also for lecturers this provides a great source, though some extra material, such as examples and explanations of different concepts and ideas behind them should be added from existing literature to make the material easier to teach and study.

JOUNI JÄRVINEN
Turku Centre for Computer Science (TUCS)
University of Turku
Turku, Finland
jouni.jarvinen@utu.fi