The First International Workshop on Rough Sets

State of the Art and Perspectives

Wojciech Ziarko

■ The First International Workshop on Rough Sets: State of the Art and Perspectives was held on 2–4 September 1992 in Kiekrz, Poland. To stimulate the discussion, the participation was limited to 40 researchers who are involved in fundamental research in rough set theory and its extensions, logic for approximate reasoning, machine learning, knowledge representation and transfer, and applications of rough set methodology. The workshop focused primarily on applications of the basic idea of the approximate definition of a set and its consequences in other areas of science and engineering. Applications discussed at the workshop included machine learning, medical diagnosis, fault detection, medical image processing, neural net training, database organization, drug research, and digital circuit design.

The First International Workshop on Rough Sets: State of the Art and Perspectives took place in Kiekrz, Poland, on 2–4 September 1992. The workshop was the first international meeting of researchers working in this relatively new area. The approximate definition of a set in terms of lower and upper bounds, as introduced in the theory of rough sets by Zdzislaw Pawlak, seems to have far-reaching consequences in theoretical mathematics and computer science as well as in empirical sciences, such as medicine and drug research. Consequently, workshop participants were a truly interdisciplinary mixture of pure theorists and down-to-earth practitioners. This mixture is what made this workshop exciting, with theoretical papers interleaved with application reports demonstrating the utility of the presented methodologies. Applications presented during the workshop included data preprocessing for neural net training, machine learning, medicine, fault detection and diagnosis, drug research, image processing, and digital circuit design. To stimulate the exchange of ideas, the workshop was limited to 40 participants.

The workshop was organized by Roman Slowinski and a group of his collaborators from the Technical University of Poznan, Poland, which is one of the main centers of applied research in rough sets in Europe. The organization was sponsored by a number of institutions, including the State Committee for Scientific Research, the Polish Science Foundation and Ministry of National Education, and the Technical University of Poznan. Fifteen selected papers will be published in a special issue of *Foundations of Computing and Decision Sciences* (Slowinski and Stefanowski 1993).

Background

The theory of rough sets has been under continuous development for over 12 years, and a fast-growing number of researchers and practitioners are interested in this methodology. The theory was originated by Pawlak in the 1970s as a result of a long-term program of fundamental research on logical properties of information systems that was carried out by him and a group of logicians from the Polish Academy of Sciences and the University of Warsaw in Poland. The methodology is concerned with the classificatory analysis of imprecise, uncertain, or incomplete information or knowledge expressed in terms of data acquired from experience. The primary notions of the theory of rough sets are the approximation space and lower and upper approximations of a set. The *approximation space* is a classification of the domain of interest into disjoint categories. The classification formally represents our knowledge about the domain; that is, the knowledge is understood here as an ability to characterize all classes of the classification, for example, in terms of features of objects belonging to the domain. Objects belonging to the same category are not distinguishable, which means that their membership status with respect to an arbitrary subset of the domain might not always be clearly definable. This fact leads to the definition of a set in terms of lower and upper approximations. The *lower approximation* is a description of the domain objects that are known with certainty to belong to the subset of interest, whereas the *upper approximation* is a description of the objects that possibly belong to the subset. Any subset defined through its lower and upper approximations is called a *rough set.* It must be emphasized that the concept of rough set should not be confused with the idea of fuzzy set because they are fundamentally different, although in some sense complementary, notions.

The specific problems addressed by the theory of rough sets are (1) representation of uncertain or imprecise knowledge, (2) empirical learning and knowledge acquisition from experience, (3) knowledge analysis, (4) analysis of conflicts, (5) evaluation of the quality of the available information with respect to its consistency and the presence or absence of repetitive data patterns, (6) iden-

tification and evaluation of data dependencies, (7) approximate pattern classification, (8) reasoning with uncertainty, and (9) information-preserving data reduction.

A number of practical applications of this approach have been developed in recent years in areas such as medicine, drug research, and process control. The recent publishing of a monograph on the theory (Pawlak 1991) and a handbook on applications (Slowinski 1992) facilitates the development of new applications. One of the primary applications of rough sets in AI is in the area of knowledge analysis and discovery in data (Ziarko 1991).

The Workshop

Workshop presentations were organized into four daily sessions, each of which contained both theoretical and application papers. The introductory plenary lecture was given by Pawlak. In his talk, he outlined the basic ideas of the rough set philosophy and emphasized the close relationship between knowledge and the ability to classify real-world objects. He also commented on current research projects involving applications of rough sets, such as data analysis, reasoning about knowledge, switching circuits, machine learning, and image processing. He suggested constructing a new kind of computer in which decision rules play the role of elementary instructions and investigating parallel implementations of rough set algorithms as future projects.

Workshop technical sessions included both theoretical and application papers. In the category of theoretical research, one could identify the following major directions: (1) research toward extending the basic rough set idea to create more accurate models of imprecise or uncertain relationships occurring in data, (2) approximate reasoning and the representation of imprecise knowledge, (3) fundamentals of machine learning in the context of rough sets, and (4) investigation of new mathematical properties stemming from the idea of rough sets.

A number of generalizations of the

basic model were proposed at the workshop. These generalizations included the replacement of the original concept of rough classification with a fuzzy similarity relation; the incorporation of probabilistic infor-

One of the primary applications of rough sets in AI is in the area of knowledge analysis and discovery in data

mation into the rough set model to better use the information hidden in data; and the generalization of the indiscernibility relation by introducing the concept of weak (partial) indiscernibility, thereby allowing for the more accurate representation of similarities among objects.

Three presentations dealt with the problem of knowledge representation and approximate reasoning in the rough set framework. A new logic for approximate reasoning, called *approximate propositional dynamic logic*, with semantics based on rough sets was proposed in the first talk, and two other presentations discussed the issues of knowledge representation in a distributed environment and the problem of multiagent knowledge transfer, particularly when the knowledge is incomplete or imprecise.

In the machine-learning papers, the main problems described were the analysis and the reduction of the combinatorial complexity of training data, the evaluation of machine-learning algorithms based on rough sets, and approximate decision making.

The newest advances in the theory of rough sets that were presented at the workshop include the investigation of topological properties of convergence of rough sets, the analysis of the computational complexity of finding the minimum nonredundant subset of attributes and the application of the Boltzman machine idea to deal with this problem, and the analysis of the relationship between the rough set model and the category theory.

Nine papers reported on applications of the rough set methodology in medicine, science, and industry. The majority of the applications were

in the medical domain, essentially dealing with the pattern-classification problem. In one application, real-life cases of patients suffering from a duodenal ulcer were classified using rough sets, and a minimal attribute–based model of these patients was built. The model was implemented on a personal computer and is currently being used in daily practice by physicians to provide indications for treatment of ulcer patients and, in particular, to assist a surgeon in determining whether surgery is required.

The objective of another medical application was to identify dominant infection factors contributing to postoperation complications. A database describing over 300 patients in terms of personal and environmental factors was used to establish relationships between the factors and the presence or absence of the infection. Rough set analysis was applied to identify and rank the most important factors.

The diagnostic classification of medical images, particularly X-ray images, using rough sets was the subject of another application study reported at the workshop. Two other talks presented the application of techniques based on rough sets to digital circuitry design and testing. The specific design problems discussed in these presentations were the minimization of Boolean functions and the derivation of the minimum set of circuit test vectors satisfying the imposed fault-detection criteria.

The other data analysis applications discussed during the workshop were the analysis of vibration symptoms for the purpose of fault detection in rolling bearings, the clustering of data in relational databases to speed up the query-processing time, and a drug research application aimed at the empirical analysis of the relationship between the structure and the activity of complex chemical compounds. An attempt to merge rough set technology with neural sets was also described.

The application involved robot training with a front end based on rough sets to eliminate irrelevant, or unimportant, sensor readings.

Summary

In summary, the workshop was well organized and offered plenty of opportunities to discuss and exchange ideas. It was an intensive and productive meeting with practically no time left for sightseeing and similar activities. It was agreed that similar workshops will be held annually in various locations in Europe and North America. I will organize the next meeting to be held in Banff, Alberta, Canada, on 12 to 15 October 1993.

Acknowledgments

Thanks are owed to Howard Hamilton, Zdzislaw Pawlak, Larry Saxton, Andrzej Skowron, and Roman Slowinski, whose detailed comments and remarks helped in the preparation of this article.
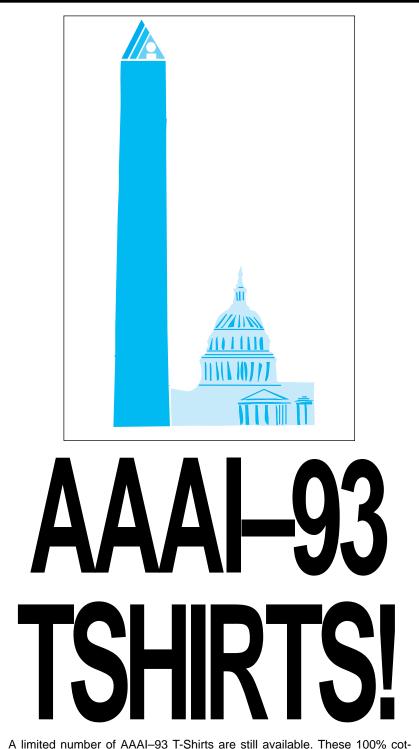
References

Pawlak, Z. 1991. *Rough Sets: Theoretical Aspects of Reasoning about Data.* Dordrecht: Kluwer Academic Publishers.

Slowinski, R., editor. 1992. *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory.* Dordrecht: Kluwer Academic Publishers.

Slowinski, R., and Stefanowski J., eds. 1993. *Foundations of Computing and Decision Sciences* 18(3–4). Forthcoming.

Ziarko, W. 1991. The Discovery, Analysis, and Representation of Data Dependencies in Databases. In *Knowledge Discovery in Databases,* eds. G. Piatetsky-Shapiro and W. J. Frawley, 177–195. Menlo Park, Calif.: AAAI Press.

Wojciech Ziarko is an associate professor in the Computer Science Department at the University of Regina, Saskatchewan, Canada. He is also vice-president of research and development at Reduct Systems Inc., a firm that is involved in developing machine-learning and knowledge-acquisition tools based on rough set methodology. He holds a Ph.D. from the Institute of Computer Science, Polish Academy of Sciences, obtained under the supervision of Zdzislaw Pawlak, the originator of the idea of rough sets. He has authored or coauthored over 50 research papers in the area of rough sets and applications, databases, and machine learning.