Rough Sets
Present State and Further Prospects

Zdzislaw Pawlak
Institute of Computer Science
Warsaw University of Technology
ul. Nowowiejska 15/19, 00-665 Warsaw, Poland

and
Institute of Theoretical and Applied Informatics
Polish Academy of Science
ul. Baltycka 5, 44 000 Gliwice, Poland

Abstract

1 Introduction

The rough set concept (cf. Pawlak (1982) is a new mathematical tool to reason about *vagueness* and *uncertainty*.

The rough set theory bears on the assumption that in order to define a set we need initially some *information* (*knowledge*) about elements of the universe - in contrast to the classical approach where the set is uniquely defined by its elements and no additional information about elements of the set is necessary. (The information about elements can be presented, for example, in a form of an *attribute-value system* called also an *information system*). Evidently, to some elements the same information can be associated and – consequently – the elements can be indiscernible in view of the available information. Thus the *indiscernibility* relation is the starting point of the rough set theory. It turns out that vagueness and uncertainty are strongly related to indiscernibility and can be defined on its basis.

2 Basic Concepts of the Rough Set Theory

Vagueness for a long time has been pursued by philosophers and logicians and in the recent years attracted attention also of AI community. The idea of a vague concept (set) is related to the so called boundary-line view, which is due to Frege (cf. Frege (1904)). The concept is vague if there some objects which can not be classified neither to the concept nor to its complement, and are the boundary-line cases.

For example, the concept of an *odd* (*even*) *number* is precise, because for each number it can be decided whether it is odd (even) or not - whereas the concept of a *beautiful women* is vague, because for some women it cannot be decided whether they are beautiful or not, (there are boundary-line cases). Difference between the lower and the upper approximation is a *boundary region* of the concept, i.e., it consists of all objects which cannot be classified with certainty to the concept or

its complement employing available knowledge. The greater the boundary region, the more vague is the concept; if the boundary region is the empty set the concept is precise.

In the rough set theory each vague concept is replaced by a pair of precise concepts called its *lower* and *upper approximations*; the lower approximation of a concept consists of all objects which *surely* belong to the concept, whereas the upper approximation of the concept consists of all objects which *possibly* belong to the concept.

Formally given any subset X of the universe U and an indiscernibility relation I, the lower and upper approximation of X are defined respectively as

$$I_*(X) = \{x \in U : I(x) \subseteq X\},$$

$$I^*(X) = \{x \in U : I(x) \cap X \neq \emptyset\},$$

where $I(x)$ denotes the set of objects indiscernible with x.

The boundary region of X is the set $BN_I(X) = I^*(X) - I_*(X)$.

If the boundary region of X is the empty set, i.e., $BN_I(X) = \emptyset$, then the set X will be called *crisp* with respect to I; in the opposite case, i.e., if $BN_I(X) \neq \emptyset$, the set X will be referred to as *rough* with respect to I.

Thus the concept of the rough set can be seen as an implementation of Frege's idea of vagueness.

Vagueness (roughness) can be characterized numerically by defining the following coefficient

$$\alpha_I(X) = \frac{|I_*(X)|}{|I^*(X)|},$$

where $|X|$ denotes the cardinality of the set X.

Obviously $0 \leq \alpha_I(X) \leq 1$. If $\alpha_I(X) = 1$, the set X is crisp with respect to I; otherwise, if $\alpha_I(X) < 1$, the set X is rough with respect to I.

A vague concept has boundary-line cases, i.e., elements of the universe which cannot be – with *certainty* – classified as elements of the concept. Hence *uncertainty* is related to the question of *membership* of elements to a set. Therefore in order to discuss the problem of uncertainty from the rough set perspective we have to define the membership function related to the rough set concept (*the rough membership function*). The rough membership function can be defined employing the indiscernibility relation I as

$$\mu_X^I(x) = \frac{|X \cap I(x)|}{|I(x)|}.$$

Obviously, $0 \leq \mu_X^I(x) \leq 1$.

The rough membership function can be used to define the approximations and the boundary region of a set, as shown below:

$$I_*(X) = \{x \in U : \mu_X^I(x) = 1\},$$

$$I^*(X) = \{x \in U : \mu_X^I(x) > 0\},$$

$$BN_I(X) = \{x \in U : 0 < \mu_X^I(x) < 1\}.$$

One can see from the above definitions that there exists a strict connection between vagueness and uncertainty in the rough set theory. As we mentioned above vagueness is related to sets, while uncertainty is related to elements of sets.

Thus approximations are necessary when speaking about vague concepts, whereas rough membership is needed when uncertain data are considered.

3 The Theory

The rough set theory have inspired a lot of theoretical research. Many authors have studied algebraic and topological properties of rough sets. Besides, a variety of logical research, directed to create logical tools to deal with approximate reasoning have been published by many authors.

The rough set concept overlaps in many aspects with many other mathematical ideas developed to deal with and vagueness and uncertainty. In particular many authors were involved in clarifying the relationship between fuzzy sets and rough sets (cf. e.g. Dubois and Parade,1992, Pawlak and Skowron,1994). Extensive study of the relation between the evidence theory and rough set theory have been revealed recently by Skowron and Grzymala-Busse (1994). The rough set philosophy in data analysis is close to statistical approach. Comparison of these two approaches can be found in Krusinska, Slowinski and Stefanowski (1992). Another aspects of statistical connections to rough sets has been considered by Wong, Ziarko and Ye (1986) and Ziarko (1993) and others. Important issue is the relationship of rough set theory to boolean reasoning, which has been deeply analyzed by Skowron and Rauszer (1992). Many authors have give attention to connections of the rough set theory and other important disciplines, like mathematical morphology, conflict theory, concurrency, Petri nets, mereology, neural networks, genetic algorithms and others.

4 Applications

After ten years of pursuing the rough set theory and its applications it is clear that this theory is of substantial importance to AI and cognitive sciences, in particular expert systems, decision support systems, machine learning, machine discovery, inductive reasoning, pattern recognition, decision tables and the like.

The rough sets approach has proved to be a very effective tool, with many successful applications to its credit. A variety of real-life applications in medicine, pharmacology, industry, engineering, control systems, social sciences, earth sciences, switching circuits, image processing and other have been successfully implemented (cf. Slowinski, 1992).

The rough set theory seems to be particularly suited to *data reduction, discovering of data dependencies, discovering data significance, discovering similarities or differences in data, discovering patterns in data, decision algorithms generation, approximate classification* and the like.

5 Further Prospects

The rough set theory have reached such a state that some kind of summary of its theoretical foundation is a must.

Besides, further development of the theory seems badly needed. The most important one seemingly is the theory of rough functions, similar to that considered in nonstandard analysis. Various approximate operations on rough functions are needed in many applications, especially in approximate (rough) control theory based on the rough set approach. Rough continuity of functions, rough stability of control systems are exemplary problems which require formulation in the framework of the rough set theory. Also appropriate formulation of complexity, which could be used to analyze control algorithms is necessary in this context.

Problems related to incomplete and distributed data seem of primary importance. Algorithms based on the rough sets approach are very well suited to parallel processing, especially when appropriate hardware could be developed. Computing machine based on the rough set concepts, seems

to be at hand. Beside practical aspects also more general look on concurrency can be gained in the framework of the rough set theory.

Closer investigation of neural networks and genetic algorithms in connection with the rough set view can contribute to better understanding the above said disciplines and lead to more efficient algorithms.

Last but not least research on rough logic seems to be very promising both theoretically and practically. The rough truth, rough consequence relation investigated by Chakraborty and Banerjee (1994) and by Lin and Liu (1994) seems to be a very good starting point to this end.

References

[1] **Black, M. (1937)**. Vagueness. *The Philosophy of Sciences.* 427-455.

[2] **Chakraborty, M.K. and Banerjee, M. (1994)**. Rough Consequence. *Bull. Pol. Acad. Sci.*, (to appear).

[3] **Dubois, D and Prade, H. (1992)**. Putting Rough Sets and Fuzzy Sets Together. In: R. Slowinski, (Ed.) *Intelligent Decision Support Systems, Handbook of Applications and Advances of the Rough Sets Theory.* KLUWER ACADEMIC PUBLISHERS, 203-232.

[4] **Fine, K. (1975)**. Vagueness, Truth and Logic. *Synthese*, 30, pp. 265-300.

[5] **Frege, G., (1904)**. Grundgesetze der Arithmentik. In: *Geach and Black (Ed.) Selections from the Philosophical Writings of Gotlob Frege*, Blackweil, Oxford 1970.

[6] **Grzymala-Busse, J. (1988)**. Knowledge Acquisition under Uncertainty - a Rough Set Approach. *Journal of Intelligent and Robotics Systems*, 1, pp. 3-16.

[7] **Krusinska, E. Slowinski, R. and Stefanowski, J. (1992)**. Discriminant Versus Rough Sets Approach to Vague Data Analysis. *Journal of Applied Statistics and Data Analysis*, 8, 43-56.

[8] **Lin, T.Y and Liu, Q (1994)**. First-Order Rough Logic I: Approximate Reasoning via Rough Sets (manuscript).

[9] **Pawlak, Z. (1982).** Rough Sets. *International Journal of Computer and Information Sciences*, 11, 341-356.

[10] **Pawlak, Z. (1981).** Pawlak. Rough Sets, Theoretical Aspects of Reasoning about Data. KLUWER ACADEMIC PUBLISHERS, 1991.

[11] **Pawlak, Z., Wong S.K.M. and Ziarko, W. (1988)**. Rough Sets: Probabilistic Versus Deterministic Approach, *Int. J. Man-Machine Studies,* 29, 81-95.

[12] **Pawlak, Z. and Skowron, A. (1994)**. Rough Membership Functions. In: Yaeger, R.R., Fedrizzi, M., and Kacprzyk, J. (Eds.), *Advances in the Dempster Shafer Theory of Evidence*, John Wiley and Sons, 251-271.

[13] **Pawlak, Z. and Slowinski, R (1994)**. Decision Analysis using Rough Sets. *Int. Trans. Opl. Res.*, 1 (1), 107-114.

[14] **Popper, K. (1959)**. The Logic of Scientific Discovery. London: Hutchinson.

[15] **Russell, B. (1923)**. Vagueness. *Australian Journal of Philosophy*, 1, 84-92.

[16] **Shafer, G. (1976)**. A Mathematical Theory of Evidence. Princeton, NJ: Princeton University Press.

[17] **Skowron, A. and Rauszer, C. (1992)**. The Discernibility Matrices and Functions in Information Systems. In : Slowinski, R. (Ed.) Intelligent Decision Support - Handbook of Advances and Applications of the Rough Set Theory. KLUWER ACADEMIC PUBLISHERS, 311-362.

[18] **Skowron, A. and Grzymala-Busse, J. (1994)**. From the Rough Set Theory to Evidence Theory. In: Yaeger, R.R., Fedrizzi, M., and Kacprzyk, J. (Eds.), *Advances in the Dempster Shafer Theory of Evidence*, John Wiley and Sons, 193-235.

[19] **Slowinski, R. (Ed.) (1992)**. Intelligent Decision Support - Handbook of Advances and Applications of the Rough Set Theory. KLUWER ACADEMIC PUBLISHERS.

[20] **Szladow, A. and Ziarko, W. (1992)**. Knowledge Based Process Control Using Rough Sets. In: R. Slowinski, ed. *Intelligent Decision Support Systems, Handbook of Applications and Advances of the Rough Sets Theory*. KLUWER ACADEMIC PUBLISHERS, 49-60.

[21] **Wong S.K.M. Ziarko, W. and Ye, R.L. (1986)**. Comparison of Rough Set and Statistical Methods in Inductive Learning. *International Journal of Man-Machine Studies*, 24, 53-72.

[22] **Zadeh, L. (1965)**. Fuzzy Sets. *Information and Control*. 338-353.

[23] **Ziarko, W. (1993)**. Variable Precision Rough Set Model, *Journal of Computer and System Sciences*, 40(1), 39-59.