Intelligent Automation and Soft Computing, Vol. 2, No. 2, pp. 95-102, 1996

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—The rough set concept is a new mathematical tool to deal with vagueness and uncertainty. It overlaps with other mathematical tools used in this area, in particular with fuzzy set theory and evidence theory. Rough set theory, however, can be viewed in its own rights, as an independent discipline. Many real-life applications of the theory have proved its practical usefulness. This article characterizes the philosophy underlying rough set theory, gives its rudiments and presents briefly some areas of application. At the end some further problems are briefly outlined.

Key Words: vagueness, uncertainty, imprecision, fuzzy sets, rough sets

1. INTRODUCTION

The rough set concept[15] is a new mathematical tool to reason about imperfect knowledge. The problem of imperfect knowledge has been tackled for a long time by philosophers, logicians and mathematicians—recently it became also interesting for computer scientists. There are many approaches to this problem. The most successful one is, no doubt, the fuzzy set theory.[31]

Rough set theory is often contrasted with fuzzy set theory. The idea of fuzzy set and rough set are not competitive, but complementary, since they refer to different aspects of imprecision. There is also a relationship between the rough set theory and Dempster-Shafer theory of evidence.[9,24] Some relations exist also between rough set theory and discriminant analysis,[10] the Boolean reasoning methods,[23] decison theory[18] and others.

Despite the relationships, rough set theory can be viewed in its own rights, as an independent discipline.

Rough set theory has found many interesting applications. The rough set approach seems to be of fundamental importance to AI and cognitive sciences, especially in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, expert systems, inductive reasoning, pattern recognition and decision support systems.

The main advantage of rough set theory is that it does not need any preliminary or additional information about data—like probability in statistics, basic probability assignment in Dempster-Shafer theory or grade of membership in fuzzy set theory. An extensive study of various mathematical models of uncertainty can be found in Grzymala-Busse.[8]

For basic ideas of the rough set theory the reader is referred to the references at the end of this article.[9,16,26,27] Many interesting applications of this approach have been presented and discussed.[12,18,25,34] The present state of rough set theory and its further perspectives are discussed in Pawlak, et al.[20]

†This work was supported by grant No. 8 S503 021 06 from State Committee for Scientific Research.

2. PHILOSOPY OF ROUGH SETS

Rough set philosophy is based on the assumption that, in contrast to classical set theory, we have some additional information (knowledge, data) about elements of a universe of discourse. Elements that exhibit the same information are indiscernible (similar) and form blocks that can be understood as elementary granules of knowledge about the universe. For example patients suffering from a certain disease, displaying the same symptoms are indiscernible and may be think of as representing a granule (disease unit) of medical knowledge. These granules are called elementary sets (concepts), and can be considered as elementary building blocks of knowledge. Elementary concepts can be combined into compound concepts, i.e., concepts that are uniquely determined in terms of elementary concepts. Any union of elementary sets is called a crisp set, and any other sets are referred to as rough (vague, imprecise).

Due to the granularity of knowledge, rough sets cannot be characterized by using available knowledge. Therefore with every rough set we associate two crisp sets, called its lower and its upper approximation. Intuitively, the lower approximation of a set consists of all elements that surely belongs to the set, whereas the upper approximation of the set constitutes of all elements that possibly belong to the set. The difference of the upper and the lower approximation is a boundary region. It consists of all elements that cannot be classified uniquely to the set or its complement, by employing available knowledge. Thus any rough set, in contrast to a crisp set, has a non-empty boundary region. For example, a concept of a beautiful painting is vague, for some paintings cannot be classified as beautiful or not, and form the boundary region of the concept—whereas the concept an odd number is crisp, because every number can be uniquely classified as odd or even, and there are no undecidable elements in this case whatsoever.

The above presented idea of vagueness, known as the "boundary-line" view, is due to Frege,[6] and it has been for a long time pursued by philosophers and logicians.[1,5,21] Thus our approach can be seen as a special case of Frege's idea.

3. BASIC CONCEPTS

In this section we give formal definitions of ideas presented above. As already mentioned, the starting point of rough set theory is the indiscernibility relation. The indiscernibility relation can be defined in purely abstract mathematical way, however, in order to provide more intuitive perspective we will use another approach—based on tabular representation of data, in a form of an attribute-value table, called also an information system. The information system is a table in which rows are labeled by elements of the universe, columns—by attributes, whereas entries of the table are values of attributes. An example of a simple information system, containing description of five toy blocks having various, size, shape and color, is shown in the table below.

	SIZE	SHAPE	COLOR
$x1$	small	triangular	red
$x2$	large	triangular	blue
$x3$	small	square	red
$x4$	small	round	yellow
$x5$	large	round	blue

In the table the universe consists of five toy blocks $x1$, $x2$, $x3$, $x4$ nd $x5$. There are three attributes in the system, SIZE, SHAPE and COLOR. The attribute SIZE has two values, "small" and "large"; the attribute SHAPE has three values, "round", "square" and "triangular"; the attribute COLOR has also three values, " red", "blue" and "yellow". The table contains data about each toy block . Each row of the table can be perceived as information (knowledge) about the corresponding toy block. For example, toy block $x3$ described by the following information: (SIZE, small), (SHAPE, square) and (COLOR, red). Each subset of attributes determines an indiscernibility relation on the set of toy blocks. For example, elements $x2$ and $x5$ are indiscernible by attributes SIZE and COLOR, elements $x1$, $x3$ and $x4$ are indiscernible in terms of attribute SIZE, and elements $x1$ and $x3$ are indiscernible by attribute COLOR.

It is easily seen that the indiscernibility relation generated by any subset of attributes is, in our example, an equivalence relation. Thus every subset of attributes determines partition of the universe into granules containing elements having the same description in terms of attribute values, i.e., elements indiscernible in terms of available information. Each such a granule can be understood as a basic building block of our knowledge about the universe.

As mentioned previously, in the rough set approach, in order to define a set we have to use information about elements of the universe.

Suppose, for example, that we would like to define set $\{x2, x3\}$ by mean of attributes SIZE and COLOR. One can easily see that this would be impossible, because elements $x1$ and $x3$ have the same description, and also elements $x2$ and $x5$ have the same description in terms of these attributes. Consequently, none of the element of the universe can be classified as surely belonging to set $\{x2, x3\}$ by employing attributes SIZE and COLOR , whereas elements $x1$, $x2$, $x3$ and $x5$ cannot be excluded as being members of the set, exploiting the information contained in the table (i.e., only element $x4$ can be definitely excluded as member of the set). This is how we arrived at the idea of approximations: the lower approximation of set $\{x2, x3\}$ is the empty set and the upper approximation of this set is $\{x1, x2, x3, x5\}$.

Now we are ready do formulate the above ideas more precisely.

Suppose we are given two finite, non-empty sets U and A, where U is the *universe*, and A is the the set of *attributes*. With every attribute $a \in A$ we associate set V_a, of its *values*, called the *domain* of a. Any subset B of A determines a binary relation $I(B)$ on U, which will be called an *indiscernibility relation*, and is defined as follows:

$$xI(B)y \text{ if and only } a(x) = a(y) \text{ for every } a \in A,$$

where $a(x)$ denotes the value of attribute a for element x.

Obviously $I(B)$ is an equivalence relation. The family of all equivalence classes of $I(B)$, i.e., partition determined by B, will be denoted by $U/I(B)$, or simple U/B; an equivalence class of $I(B)$, i.e., block of the partition U/B, containing x will be denoted by $B(x)$. If (x, y) belongs to $I(B)$ we say that x and y are *indiscernible* with respect to B.

Let us define now two following operations on sets

$$B_*(X) = \{x \in U: B(x) \subseteq X\}$$

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

assigning to every subset X of the universe U two sets $B_*(X)$ and $B^*(X)$ called the *B-lower* and the *B-upper approximation* of X, respectively. The set

$$BN_B(X) = B^*(X) - B_*(X)$$

will be referred to as the *B-boundary region* of X.

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

The boundary region of a set consists of all elements of the universe which cannot be classified to the set or its complement, on the basis of their features. This idea can be also characterized numerically by a coefficient, called the *accuracy of approximation* and defined as follows:

$$\alpha_B(X) = \frac{|B_*(X)|}{|B^*(X)|}$$

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

Obviously $0 \le \alpha_B(X) \le 1$. If $\alpha_B(X) = 1$, the set X is crisp with respect to B; otherwise i.e., if $\alpha_B(X) < 1$, the set X is rough with respect to B.

It is worthwhile to notice[15] that approximations are interior and closure operations in a topology generated by the indiscernibility relation. Consequently, the above definition of the rough set has its roots in topology. One can give also several other equivalent definitions of rough sets, however we will exclude this issue here.

Sets are generally defined by means of a membership function. In the presented approach instead of the membership function we used approximations. However, rough sets can be also defined using the concept of membership, by defining the *rough membership function* as shown below,

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

Obviously

$$\mu_X^B(x) \in [0,1]$$

It might seem that this is a fuzzy membership function, but this is not the case, as it is clearly seen from the following properties[17]

1.) $\mu_{U-X}^B(x) = 1 - \mu_X^B(x)$ for any $x \in U$

2.) $\mu_{X \cup Y}(x) \ge \max\left(\mu_X^B(x), \mu_Y^B(x)\right)$ for any $x \in U$

3.) $\mu_{X \cap Y}^B(x) \le \min\left(\mu_X^B(x), \mu_Y^B(x)\right)$ for any $x \in U$

Besides, the rough membership, in contrast to fuzzy membership, has obvious probabilistic flavor, and can be interpreted as a kind of conditional probability.

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

$$B_*(X) = \left\{x \in U: \mu_X^B(x) = 1\right\}$$
$$B^*(X) = \left\{x \in U: \mu_X^B(x) > 0\right\}$$
$$BN_B(X) = \left\{x \in U: 0 < \mu_X^B(x) < 1\right\}$$

Thus, we have two ways of defining rough sets, by using topological and probabilistic methods. However, one can show[17] that both definitions are not equivalent. This should not be a surprise, since the first definition refers to sets, and expresses the impossibility of defining a set by means of certain tools, while the second one is connected with elements of a set, and expresses our inability to classify elements to some concepts. Hence both definitions refer to different aspects of imperfect knowledge, the first one to vagueness of concepts, while the second—to uncertainty of elements.

4. THE THEORY

The rough set theory has 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.[4,17] Extensive study of the relation between the evidence theory and rough set theory has been revealed recently in Skowron and Grzymala-Busse (1994).[24] 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).[10] Another aspects of statisticals and rough sets has been considered in Wong, Ziarko, and Ye (1986)[30]and Ziarko (1993).[34] and others. Important issue is the relationship of rough set theory to boolean reasoning, which has been deeply analyzed in Skowron and Rauszer (1992).[23] Many authors have given attention to connections of rough set theory and other important disciplines, like mathematical morphology, conflict theory, concurrency, Petri nets, mereology, neural networks, genetic algorithms and others.

5. APPLICATIONS

After several years of pursuing 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.[25]

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.

The rough set methodology has found many real-life applications in medical data analysis, finance, voice recognition, image processing and others. [12,13,25,26,29,32,34]

The proposed method has many important advantages. Some of them are listed below.

- Provides efficient algorithms for finding hidden patterns in data
- Finds minimal sets of data (data reduction)
- Evaluates significance of data
- Generates minimal sets of decision rules from data
- It is easy to understand
- Offers straightforward interpretation of obtained results
- Most algorithms based on the rough set theory are particularly suited for parallel processing, but in order to exploit this feature fully, a new hardware is necessary.

Many other fields of applications of the rough set theory are being extensively explored.

6. FURTHER PROSPECTS

Rough set theory has reached a certain degree of maturity. More than a thousand papers have been hitherto published on rough set theory and its applications. Despite many important theoretical contributions and extensions of the original model many essential theoretical problems still remain open and besides, further development of the theory seems badly needed.

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.

Research on rough logic seems to be very promising both theoretically and practically. The rough truth, rough consequence relation investigated in Chakrabory and Banerjee[2] and in Lin and Liu[11] seems to be a very good starting point to this end.

Theory of rough functions, similar to that considered in nonstandard analysis, is seemingly also very importent issue. 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.

In summary

- rough logic, based on the concept rough truth seems to be a very important issue
- theory of rough relation and rough function is necessary in many applications
- comparison with many other theories and tools is of primary significance

Besides, some practical problems related with application of rough sets in many domains are of great importance. Particularly

- efficient and widely accessible software is necessary to further development of various applications
- development of rough set computer seems to be a must in order to pursue many new applications
- last but not least "rough control" seems to be a very promising area of application of the rough set concept

REFERENCES

1. Black, M. "Vagueness". *The Philosophy of Sciences*, (1937), pp. 427-455.
2. Chakraborty, M.K. and M. Banerjee. "Rough Consequence." *Bull. Pol. Acad. Sci.*, (to appear).
3. Czogala, E.., A. Mrozek, and Z. Pawlak. "The Idea of Rough-Fuzzy Controller." *International Journal of Fuzzy Sets and Systems*, 72 (1995), pp. 61-63.
4. Dubois, D., and H. Prade. "Putting Rough Sets and Fuzzy Sets Together." In *Intelligent Decision Support - Handbook of Advances and Applications of the Rough Set Theory*, R. Slowinski ed., Kluwer Academic Publishers, Dordrecht, Boston, London (1992), pp. 203-232.
5. Fine, K. "Vagueness, Truth and Logic." *Synthese*, 30 (1975), pp. 265-300.
6. Frege, G. "Grundgesetze der Arithmentik." In *Geach and Black (Ed.) Selections from the Philosophical Writings of Gotlob Frege*, 2, Blackweil, Oxford (1970).
7. Grzymala-Busse, J. "Knowledge Acquisition under Uncertainty - a Rough Set Approach". *Journal of Intelligent and Robotics Systems*, 1 (1988), pp. 3-16.
8. Grzymala-Busse, J. *Managing Uncertainty in Expert Systems*. Kluwer Academic Publishers, Dordrecht, Boston, London (1991).
9. Grzymala-Busse, J. "Rough Sets." *Advances in Imaging and Electrons Physics*, 94, (1995).
10. Krusinska, E.., R. Slowinski, and J. Stefanowski. "Discriminant versus Rough Set Approach to Vague Data Analysis."

Journal of Applied Statistics and Data Analysis, 8 (1992), pp. 43-56.

11. Lin, T.Y., and Q. Liu. "First-Order Rough Logic I: Approximate Reasoning via Rough Sets" (manuscript, 1994).
12. Lin, T.Y., and A.M. Wilderberg, eds. *Soft Computing*, Proceedings of the Third International Workshop on Rough Sets and Soft Computing (RSSC '94), November 10-12, 1994, San Jose State University, San Jose, California, USA (1994).
13. Lin, T.Y., ed. CSC '95, 23rd Annual Computer Science Conference on Rough Sets and Database Mining, Conference Proceedings, March 2, San Jose State University, San Jose, California, USA (1995).
14. Munakata, T. "Rough Control: A perspective". In *CSC'95, 23rd Annual Computer Science Conference on Rough Sets and Database Mining*, Conference Proceedings, March 2, San Jose State University, San Jose, California, USA (1995).
15. Pawlak, Z. "Rough Sets." *International Journal of Computer and Information Sciences*,11 (1982), pp. 341-356.
16. Pawlak, Z. *Rough Sets, Theoretical Aspects of Reasoning* about Data, Kluwer Academic Publishers, Dordrecht, Boston, London (1991).
17. Pawlak, Z., and A. Skowron. "Rough Membership Functions." In *Advances* in *the Dempster-Shafer Theory of Evidence*, R.R. Yeager, M. Fedrizzi, J. Kcprzyk, eds., John Wiley and Sons, New York (1994), pp. 251-271.
18. Pawlak, Z., and R. Slowinski. "Decision Analysis using Rough Sets." *International Transactions on Operational Research*}, 1 (1994), pp. 107-104.
19. Pawlak, Z., S.K.M. Wong, W. Ziarko. "Rough Sets: Probabilistic versus Deterministic Approach", *International Journal of Man Machine Studies*, 29 (1988), pp. 81-85.
20. Pawlak, Z., J. Grzymala-Busse, R. Slowinski, and W. Ziarko. "Rough Sets." *Communication of the ACM* (1995).
21. Russell, B. "Vagueness." *Australian Journal of Philosophy*, 1 (1923), pp. 84-92.
22. Shafer, G. A Mathematical Theory of Evidence. Princeton, NY: Princeton University Press (1976).
23. Skowron, A., and C. Rauszer. "The Discernibility Matrices and Functions in Information Systems." In *Intelligent Decision Support - Handbook of Advances and Applications of the Rough Set Theory* R. Slowinski, ed., Kluwer Academic Publishers, Dordrecht, Boston, London (1992), pp. 311-369.
24. Skowron, A., and J. Grzymala-Busse. "From Rough Set Theory to Evidence Theory." In *Advances in the Dempster-Shafer Theory of Evidence* R.R. Yeager, M. Fedrizzi, J. Kcprzyk, eds., John Wiley and Sons, New York (1994), pp. 251-271, pp. 193-235.
25. Slowinski, R., ed. *Intelligent Decision Support - Handbook of Advances and Applications of the Rough Set Theory*, Kluwer Academic Publishers, Dordrecht,Boston, London (1992).
26. Slowinski, R., and C. Zopounidis. "Application of the Rough Set Approach To Evaluation of Bankruptcy Risk." *Intelligent Systems in Accounting, Finance and Management*, 4 (1995), pp. 27-41.
27. Slowinski, R. "Rough Set Approach to Decision Analysis." *AI Expert*, 10 (1995), pp. 18-25.
28. Szladow, A., and W. Ziarko. "Rough Sets: Working with Imperfect Data." *AI Expert*, 8 (1993), pp. 36-41.
29. Woolery, L.K., and J. Grzymala-Busse. "Machine Learning for an Expert System to Predict Preterm Birth Risk." *Journal of the American Medical Informatics Association* 1 (1994), pp. 439-446.
30. Wong, S.K.M., W. Ziarko, and R.L. Ye. "Comparison of Rough Set and Statistical Methods in Inductive Learning." *International Journal of Man-Machine Studies*, 24 (1986), pp. 53-72.
31. Zadeh, L. "Fuzzy Sets." *Information and Control*. (1965), pp. 338-353.
32. Ziarko,W., R. Golan, and D. Edwards. "An Application of DATALOGIC/R Knowledge Discovery Tool to Identify Strong Predicative Rules in Stock Market Data." In *roceedings of AAAI Workshop on Knowledge Discovery in Databases*, Washington DC, (in press).
33. Ziarko, W. "Variable Precision Rough Set Model." *Journal of Computer and System Sciences*, 40 (1993), pp. 39-59.
34. Ziarko, W., ed. *Rough Sets, Fuzzy Sets and Knowledge Discovery*. Proceeding of the International Workshop on Rough sets and Knowledge Discovery (RSKD'93), Banff, Alberta, Canada , Springer Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest (1993).