ROUGH SETS PRESENT
STATE AND FURTHER PROSPECTS

by

Zdzisław Pawlak

ICS Research Report 32/95

Institute of Computer Science

Nowowiejska 15/19, 00-665 Warsaw, Poland
ROUGH SETS
PRESENT STATE AND FURTHER PROSPECTS *

Zdzislaw Pawlak
Institute of Computer Science, Warsaw University of Technology
Nowowiejska 15/19, 00-665 Warsaw, Poland
e-mail: zpw@ii.pw.edu.pl

and
Institute of Theoretical and Applied Informatics
Polish Academy of Sciences
ul. Bałtycka 5, 44-100 Gliwice, Poland

Abstract

The rough set theory is a new mathematical approach to vagueness and uncertainty. To some extent it overlaps with some other mathematical tools developed to deal with imperfect knowledge, in particular with fuzzy set theory and evidence theory — nevertheless the rough set theory can be viewed in its own rights, as an independent discipline. Many real-life applications of the theory have proved its practical usefulness. The paper characterizes the philosophy underlying the rough set theory, gives its rudiments and discusses briefly some areas of applications. At the end some further problems are briefly outlined.

Keywords: Vagueness, Uncertainty, Imprecision, Fuzzy sets, Rough sets

1 Introduction

The rough set theory proposed by the author in 1982 [21] is a new mathematical tool to reason about vagueness and uncertainty.

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 [8]. The concept is vague if there are some objects which cannot be classified neither to the concept nor to its complement, and are the boundary-line cases.

*This work was supported by grant No. 8 SS03 021 06 from State Committee for Scientific Research.
The rough sets philosophy bears on the assumption that knowledge has granular structure. The granularity of knowledge is caused by the fact that often some objects of interest cannot be discern and may appear as the same (or similar). Consider, for example patients suffering from a certain disease. Suppose that with every patient a set of data (like body temperature, blood pressure, etc.) characterizing his/her health status is associated. Thus patients displaying the same symptoms are indiscernible (similar) in view of the available knowledge about them. The indiscernibility relation generated in this way is the mathematical basis of the rough set theory.

Any set of all indiscernible (similar) objects is called elementary set, and form basic granule (atom) of knowledge about the universe. Any union of some elementary sets is referred to as crisp (precise) set – otherwise a set is rough (imprecise, vague).

Consequently each rough set has boundary-line examples, i.e. objects which cannot be with certainty classified, employing the available knowledge, as members of the set or its complement.

In the proposed approach we assume that any vague concept is characterized by pair of precise concepts – called the lower and the upper approximation of the vague concept. The lower approximation consists of all objects which surely belong to the concept and the upper approximation contain all objects which possible belong to the concept. Obviously the difference between the upper and the lower approximation constitute the boundary region of the vague concept. Approximations are two basic operations in the rough set theory.

The rough set concept overlaps – to some extent – with many other mathematical tools developed to deal with imperfect knowledge, in particular with the fuzzy set theory and the theory of evidence.

Basically the idea of fuzzy set and rough set are not competitive, since they refer to different aspects of imprecision, and consequently are meant to be used in different areas. The relationship between these two approaches has been studied by many authors. Extensive discussion of this topic
can be also found in [23].

The connection between the rough set theory and Dempster-Shafer theory of evidence has been investigated in [31]. It turned out that the plausibility and belief functions, used in the evidence theory, can be expressed in the rough set theory by employing the lower and the upper approximations. Hence the theory of evidence can be embedded in the rough set theory.

2 Illustrative Example

The above presented ideas can be illustrated by the following example. Suppose we are given data table – called also attribute-value table or information system – containing data about 6 patients, as shown below.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Headache</th>
<th>Muscle-pain</th>
<th>Temperature</th>
<th>Flu</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>no</td>
<td>yes</td>
<td>high</td>
<td>yes</td>
</tr>
<tr>
<td>p2</td>
<td>yes</td>
<td>no</td>
<td>high</td>
<td>yes</td>
</tr>
<tr>
<td>p3</td>
<td>yes</td>
<td>yes</td>
<td>very high</td>
<td>yes</td>
</tr>
<tr>
<td>p4</td>
<td>no</td>
<td>yes</td>
<td>normal</td>
<td>no</td>
</tr>
<tr>
<td>p5</td>
<td>yes</td>
<td>no</td>
<td>high</td>
<td>no</td>
</tr>
<tr>
<td>p6</td>
<td>no</td>
<td>yes</td>
<td>very high</td>
<td>yes</td>
</tr>
</tbody>
</table>

Tab.1

Columns of the table are labelled by attributes (symptoms) and rows by objects (patients), whereas entries of the table are attribute values. Thus each row of the table contains knowledge (information) about specific patient. For example patient p2 is characterized in the table by the following attribute-value set

\{
(Headache, yes), (Muscle-pain, no), (Temperature, high), (Flu, yes)\}.

In the table patients p2, p3 and p5 are indiscernible with respect to the attribute Headache, patients p3 and p6 are indiscernible with respect to attributes Muscle-pain and Flu, and patients
p2 and p5 are indiscernible with respect to attributes Headache, Muscle-pain and Temperature. Hence, for example, the attribute Headache generates two elementary sets \{p2, p3, p5\} and \{p1, p4, p6\}, whereas the attributes Headache and Muscle-pain form the following elementary sets, \{p1, p4, p6\}, \{p2, p5\} and \{p3\}. Similarly one can define elementary set generated by any subset of attributes.

Because patient p2 has flu, whereas patient p5 does not, and they are indiscernible with respect to the attributes Headache, Muscle-pain and Temperature, thus flu cannot be characterized in terms of attributes Headache, Muscle-pain and Temperature. Hence p2 and p5 are the boundary-line cases, which cannot be properly classified in view of the available knowledge. The remaining patients p1, p3 and p6 display symptoms which enable us to classify them with certainty as having flu, patients p1 and p5 cannot be excluded as having flu and patient p4 for sure has not flu, in view of the displayed symptoms. Thus the lower approximation of the set of patients having flu is the set \{p1, p3, p6\} and the upper approximation of this set is the set \{p1, p2, p3, p5, p6\}, where as the boundary-line cases are patients p2 and p5. Similarly p4 has not flu and p2, p5 cannot be excluded as having flu, thus the lower approximation of this concept is the set \{p4\} whereas – the upper approximation is the set \{p2, p4, p5\} and the boundary region of the concept "not flu" is the set \{p2, p5\} the same

3 Basic Concepts

The above discussed ideas can be formulated more precisely as shown in what follows.

Let \(U\) be a non-empty set called the universe and let \(I\) be a binary relation over \(U\), called the indiscernibility relation. In what follows, for the sake of simplicity, we assume that \(I\) is an equivalence relation and \(I(x)\) will denote the equivalence class containing \(x\).

Let \(X \subseteq U\) be a subset of the universe. By the lower approximation of \(X\) with respect to \(I\), or
briefly the lower approximation of $X$, when $I$ is understood, we understand the set

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

and similarly the upper approximation of $X$ with respect to $I$, is the set

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

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 $I$.

Let us remark that the lower and the upper approximations are in fact interior and closure in a certain topology generated by the indiscernibility relation [22].

The greater the boundary region of a set, the more "rough" (vague) is the set. The above idea can be expressed 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 a 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

$$\mu^I_X(x) = \frac{|X \cap I(x)|}{|I(x)|}.$$
Obviously, $0 \leq \mu^I_X(x) \leq 1$.

The rough membership function can be interpreted as a kind of conditional probability, and is not assumed, but computed from data about objects of the universe.

Using the rough membership function one can define the rough set as follows: a set $X$ is rough with respect to $I$ if $\mu^I_X(x) < 1$ for every $x \in X$, otherwise, i.e. if $\mu^I_X(x) = 1$ for every $x \in X$, the set $X$ is crisp with respect to $I$. It turns out that the definition of the rough set concept employing approximations and that based on the rough membership function are not equivalent. For details the reader is referred to [23].

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

$$L_I(X) = \{x \in U : \mu^I_X(x) = 1\},$$

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

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

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, and has topological flavor, while uncertainty is related to elements of a set and it has probabilistic nature.

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

4 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 vagueness and uncertainty. In particular many authors were involved in clarifying the relationship between fuzzy sets and rough sets [2,23]. Extensive study of the relation between the evidence theory and rough set theory have been revealed recently by [31]. The rough set philosophy in data analysis is close to statistical approach. Comparison of these two approaches can be found in [13]. Another aspects of statistical connections to rough sets has been considered by [42,47] and others. Important issue is the relationship of rough set theory to boolean reasoning, which has been deeply analyzed by [30]. Many authors have given 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.

5 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. Some of them are listed in the References, especially in [32,47,15].

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.
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 and discrete dynamical systems. Many basic properties of functions such as rough continuity, rough limits, rough derivative, rough integral and rough stability are exemplary problems which require formulation in the framework of the rough function theory. This approach can also contribute to qualitative reasoning methods already studied recently in physics and AI.

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 theory 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 [2,16] seems to be a very good starting point to this end.
Conclusion

It seems that the rough set theory has reached a certain degree of maturity both from theoretical and practical points of view.

It has inspired wide spectrum of theoretical research in various areas and also 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.

Recent research on rough control have shown a new very promising area of applications of the rough set theory.

References


15. Lin, T. Y., ed., The Third International Workshop on Rough Sets and Soft Computing Pro-
17. Munakata, T., "Rough Control a Perspective". In 23rd Annual Computer Science Conference
Workshop on Rough Sets and Database Mining, (CSC'95), Conference Proceedings, Lin, T. Y., ed.,
March 2, 1995, San Jose State University, San Jose, California, 1995.
18. Mrozek, A., "Rough Sets in Computer Implementation of Rule-Based Control of Industrial
Processes". In Intelligent Decision Support Systems, Handbook of Applications and Advances of the
Diagnostics by means of the Rough Sets Theory". In Intelligent Decision Support Systems, Hand-
21. Pawlak, Z., "Rough Sets", International Journal of Computer and Information Sciences, 11,
22. Pawlak, Z., Rough Sets, Theoretical Aspects of Reasoning about Data, Kluwer Academic Pub-
23. Pawlak, Z., Skowron, A., "Rough Membership Functions". In Advances in the Dempster Shafer
24. Pawlak, Z., Slowinski, R., "Decision Analysis using Rough Sets", International Translations on
30. Skowron, A., Rauszer, C., "The Discernibility Matrices and Functions in Information Systems".


36. Slowinski, K., "Rough Classification of HSV Patients". In Intelligent Decision Support Systems,


36. Slowinski, K., "Rough Classification of HSV Patients". In Intelligent Decision Support Systems,


Recently published Research Reports of the Institute of Computer Science, W.U.T.


