Why Rough Sets? *

Zdzisław Pawlak
Institute of Computer Science, Warsaw University of Technology
ul. Nowowiejska 15/19, 00665 Warsaw, Poland
zpw@ii.pw.edu.pl and
Institute of Theoretical and Applied Informatics
Polish Academy of Sciences
ul. Baltycka 5, 44 000 Gliwice, Poland

Abstract

The problem of imperfect knowledge has been tackled for a long time by philosophers, logicians and mathematicians. Recently it became also a crucial issue for computer scientists, particularly in the area of artificial intelligence. There are many approaches to the problem of how to understand and manipulate the imperfect knowledge. The most successful one is, no doubt, fuzzy set theory proposed by Zadeh.

Rough set theory is another attempt to this problem. The theory has attracted attention of many researchers and practitioners all over the world, who contributed essentially to its development and applications.

Rough set theory overlaps with many other theories, especially with fuzzy set theory, evidence theory and Boolean reasoning methods – nevertheless it can be viewed in its own rights, as an independent, complementary, and not competing discipline.

1 Introduction

The rough set philosophy is founded on the assumption that with every object of the universe of discourse we associate some information (data, knowledge). E.g., if objects are patients suffering from a certain disease, symptoms of the disease form information about patients.

All patients revealing the same symptoms are indiscernible (similar) in view of the available information and form blocks, which can be understood as elementary granules of knowledge about patients (or types of patients). These granules are called *elementary sets* or *concepts*, and can be considered as elementary building blocks of our knowledge. Elementary concepts can be combined into compound concepts, i.e. concepts that are uniquely defined 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*).

Consequently each rough set has boundary-line cases, i.e., objects which cannot be with certainty classified as members of the set or of its complement. Obviously crisp sets have no boundary-line elements at all. That means that boundary-line cases cannot be properly classified by employing the available knowledge.

Thus, the assumption that objects can be "seen" only through the information available about them leads to the view that knowledge has granular structure. Due to the granularity of knowledge some objects of interest cannot be discerned and appear as the same (or similar). As, a consequence vague concepts, in contrast to precise ones, cannot be characterized in terms of information about their elements. Therefore in the proposed approach we assume that any vague concept is replaced by a 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 contains 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 rough set theory.

Rough set theory overlaps to a certain degree many other mathematical theories. Particularly interesting is the relationship with fuzzy set theory and Dempster-Shafer theory of evidence. The concepts of rough set and fuzzy set are different since they refer to various aspects of imprecision [12] whereas the

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

connection with theory of evidence is more substantial [17]. Besides, rough set theory is related to discriminant analysis [6], Boolean reasoning methods [16] and others. The relationship between rough set theory and decision analysis is presented in [13,19].

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

The rough set approach seems to be of fundamental importance to AI and congnitive sicences, especially in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, expert systems, inductive reasoning and pattern recognition.

The main advantage of rough set theory is that it does not need any preliminary or additional information about data – like probability in statistics, or basic probability assignment in Dempster-Shafer theory and grade of membership or the value of possibility in fuzzy set theory.

The rough set theory has been successfully applied in many real-life problems in medicine, pharmacology, engineering, banking, financial and market analysis and others.

2 An Example

Data are often presented as *information tables*, column of which are labelled by *attributes*, rows by *objects* and entries of the table are *attribute-values*. Simple illustrative example of information table is shown below.

Patient	Headache	Muscle-pain	Temp.	Flu
p1	no	yes	high	yes
p2	yes	no	high	yes
p3	yes	yes	v. high	yes
p4	no	yes	normal	no
p5	yes	no	high	no
p6	no	yes	v. high	yes

Table 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 can be seen as 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),

which form information about the patient.

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 sets 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 p2 and p5 cannot be excluded as having flu and patient p4 for sure does not have 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}, whereas the boundary-line cases are patients p2 and p5. Similarly p4 does not have flu and p2, p5 cannot be excludes 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 as in the previous case.

3 Formal Definition of Approximations

Now we present above concepts more formally.

Suppose we are given two finite, non-empty sets U and A, where U is the *universe*, and A – a set *attributes*. With every attribute $a \in A$ we associate a 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:

$xI(B)y$ if and only if $a(x) = a(y)$ 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 where

$$POS_C(D) = \bigcup_{X \in U/I(D)} C_*(X).$$

The expression $POS_C(D)$, called a *positive region* of the partition U/D with respect to C, is the set of all elements of U that can be uniquely classified to blocks of the partition U/D, by means of C.

Thus the coefficient k expresses the ratio of all elements of the universe, which can be properly classified to blocks of the partition U/D, employing attributes C. Notice that for $k = 1$ we get the previous definition of total dependency.

For dependency {Headache, Muscle-pain, Temperature} $\Rightarrow$ {Flu} we get $k = 4/6 = 2/3$, because four out of six patients can be uniquely classified as having flu or not, employing attributes Headache, Muscle-pain and Temperature.

If we were interested in how exactly patients can be diagnosed using only the attribute Temperature, that is – in the degree of the dependence {Temperature} $\Rightarrow$ {Flu}, we would get $k = 3/6 = 1/2$, since in this case only three patients $p3, p4$ and $p6$ out of six can be uniquely classified as having flu. In contrast to the previous case patient $p4$ cannot be classified now as having flu or not. Hence the single attribute Temperature offers worse classification than the whole set of attributes Headache, Muscle-pain and Temperature. It is interesting to observe that neither Headache nor Muscle-pain can be used to recognize flu, because for both dependencies {Headache} $\Rightarrow$ {Flu} and {Muscle-pain} $\Rightarrow$ {Flu} we have $k = 0$.

Summing up: D is *totally* (*partially*) dependent on C, if *all* (*some*) elements of the universe U can be uniquely classified to blocks of the partition U/D, employing C.

5 Reduction of Attributes

We often face a question whether we can remove some data from an information table preserving its basic properties, that is – whether a table contains some superfluous data. For example, it is easily seen that if we drop in Table 1 either the attribute Headache or Muscle-pain we get the data set which is equivalent to the original one, in regard to approximations and dependencies. That is we get in this case the same accuracy of approximation and degree of dependencies as in the original table, however using smaller set of attributes.

This concept can be formulated more precisiely as follows. Let D depends on C. A minimal subset C' of C, such that D depends on C' is called a *reduct* of C.

Thus a reduct is a set of attributes that preserves the dependence. It means that a reduct is a minimal subset of attributes that enables the same classification of elements of the universe as the whole set of attributes.

Obviously a set of attributes may have more then one reduct. Intersection of all reducts is called the *core*. The core in Table 1 is the attribute Temperature. Thus the core is, in a certain sense, the set of the most important attributes, that cannot be eliminated from the information table without changing its dependencies.

6 Decison Rules

Sometimes we distinguish in an information table two classes of attributes, called *condition* and *decision* (*actions*) attributes. For example in Table 1 attributes Headache, Muscle-pain and Temperature can be considered as condition attributes, whereas the attribute Flu – as a decision attribute.

Each row of a decision table detremines a *decision rule*, which specifies *decisions* (*actions*) that should be taken when conditions pointed out by *condition* attributes are satisfied. For example in Table 1 the condition (Headache, no), (Muscle-pain, yes), (Temperature, high) determines uniquely the decision (Flu, yes). Decision rules 2) and 5) in Table 1 have the same conditions by different decisions. Such rules are called *inconsistent* (*nondeterministic, conflicting*); otherwise the rules are referred to as *consistent* (*deterministic, nonconflicting*). Sometimes consistent decision rules are called *sure* rules, and inconsistent rules are called *possible* rules. Decision tables containing inconsistent decision rules are called (inconsistent);otherwise the table is *consistent* . Similar terminology applies to decision tables.

The number of consistent rules to all rules in a decision table can be used as *consistency measure* of the decision table, and will be denoted by $\gamma(C, D)$, where C and D are condition and decision attributes respectively. Thus if $\gamma(C, D) = 1$ the decision table is consistent and if $\gamma(C, D) \neq 1$ the decision table is inconsistent. Notice, that $\gamma(C, D) = \frac{POS_c(D)}{|U|}$.

Decision rules are often presented as implications and are called "*if... ,then...*" rules. For example rule 1) in Table 1 can be presented as implication

if (Headache, no) and (Muscle-pain, yes) and (Temperature, high) then (Flu, yes).

Hence a decison table can be wieved as a set of decison rules. With every total or partial dependency we can associate a set of decison rules which uniqly determines the dependency. Beside reducing condition partition U/B, containing x will be denoted by $B(x)$.

If (x, y) belongs to $I(B)$ we will say that x and y are *B-indiscernible*. Equivalence classes of the relation $I(B)$ (or blocks of the partition U/B) are refereed to as *B-elementary sets*.

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 is *crisp* (*exact*) with respect to B; in the opposite case, i.e., if $BN_B(X) \neq \emptyset$, the set X is *rough* (*inexact*) with respect to B.

Rough set can be also characterized numerically by the following coefficient

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

called *accuracy of approximation*, where $|X|$ denotes the cardinality of X. Obviously $0 \leq \alpha_B(X) \leq 1$. If $\alpha_B(X) = 1$, X is *crisp* with respect to B (X is *precise* with respect to B), and otherwise, if $\alpha_B(X) < 1$, X is *rough* with respect to B (X is *vague* with respect to B).

Let us depict above definitions by examples referring to Table 1. Consider the concept "flu", i.e., the set $X = \{p1, p2, p3, p6\}$ and the set of attributes B = {Headache, Muscle-pain, Temperature}. Concept "flu" is roughly B-definable, because $B_*(X) = \{p1, p3, p6\} \neq \emptyset$ and $B^*(X) = \{p1, p2, p3, p5, p6\} \neq U$. For this case we get $\alpha_B(\text{"}flu\text{"}) = 3/5$. It means that the concept "flu" can be characterized partially employing symptoms, Headache, Muscle-pain and Temperature. Taking only one symptom B = {Headache} we get $B_*(X) = \emptyset$ and $B^*(X) = U$, which means that the concept "flu" is totally indefinable in terms of attribute Headache, i.e., this attribute is not characteristic for flu whatsoever. However, taking single attribute B = {Temperature} we get $B_*(X) = \{p3, p6\}$ and $B^*(X) = \{p1, p2, p3, p5, p6\}$, thus the concept "flu" is again roughly definable, but in this case we obtain $\alpha_B(X) = 2/5$, which means that the single symptom Temperature is less characteristic for flu, than the whole set of symptoms, and patient p1 cannot be now classified as having flu in this case.

Rough sets can be also defined using a *rough membership function*, defined as

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

Obviously

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

Value of the membership function $\mu_X(x)$ is kind of conditional probability, and can be interpreted as a degree of *certainty* to which x belongs to X (or $1 - \mu_X(x)$, as a degree of *uncertainty*).

4 Dependency of Attributes

Another important issue in data analysis is discovering dependencies between attributes. Intuitively, a set of attributes D depends totally on a set of attributes C, denoted $C \Rightarrow D$, if all values of attributes from D are uniquely determined by values of attributes from C. In other words, D depends totally on C, if there exists a functional dependency between values of D and C. In Table 1 there are not total dependencies whatsoever. If in Table 1, the value of the attribute Temperature for patient $p5$ were "no" instead of "high", there would be a total dependency {Temperature} $\Rightarrow$ {Flu}, because to each value of the attribute Temperature there would correspond unique value of the attribute Flu.

Formally dependency can be defined in the following way. Let D and C be subsets of A. We say that *D depends totally* on C, if and only if $I(C) \subseteq I(D)$. That means that the partition generated by C is finer than the partition generated by D. Notice, that the concept of dependency discussed above corresponds to that considered in relational databases.

We would need also a more general concept of dependency of attributes, called a *partial dependency* of attributes. Let us first depict the idea by example, referring to Table 1. In this table, for example, the attribute Temperature determines uniquely only some values of the attribute Flu. That is, (Temperature, very high) implies (Flu, yes), similarly (Temperature, normal) implies (Flu, no), but (Temperature, high) does not imply always (Flu, yes). Thus the partial dependency means that only some values of D are determined by values of C.

Formally, the above idea can be formulated as follows. Let D and C be subsets of A. We say that *D depends in degree $k, 0 \leq k \leq 1$*, on C, denoted $C \Rightarrow_k D$, if

$$k = \frac{|POS_C(D)|}{|U|},$$

attributes in an a decison table we can also reduce attributes in each decison rule obtaing minmal decison rules. For example with the partial dependecy

{Headache, Muscle-pain, Temperature} $\Rightarrow$ {Flu}

we can associate the folowing minimal set of decison rules

If (Temperature, normal), then (Flu, yes),

If (Headache, no) and (Muscle-pain, yes) or
(Muscle-pain, yes) and (Temperature, high)
(Temperature, very high),
then (Flu, yes).

7 Applications and Advantages

The rough set methodology has found many real-life applications.

There are many applications in medicine, pharmacology, banking, and market research. Very interesting results have been also obtained in speaker independent speech recognition, and acoustics. The rough set approach seems also important for various engineering applications, like diagnosis of machines using vibroacoustics, symptoms (noise, vibrations), material sciences and process control. Application in linguistics and environment, databses are other important domains.

More about applications of the rough set theory can be found in [7,18,22]. Besides, many other fields of application, e.g., time series analysis, image processing and character recognition, are being extensively explored.

Application of rough sets requires a suitable software. Many software systems for workstations and personal computers based on rough set theory have been developed. Some of them are available commercially.

The theory 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.

8 Further Research

More than 1000 papers have been published on rough set theory and its applications till now.

Although rough set theory has many achierements to its credit, nevertheless several theoretical and practical problems require further attention.

Especially important is widly accessible efficient software development for rough set based data analysis, particularly for large collections of data analysis.

Despite of many valuable methods of efficient, optimal decision rule generation methods from data, developed in recent years based on rough set theory – more research here is needed, particularly, when quantitative attributes are involved. In this context also further discretization methods for quantitative attribute values are badly needed. Comparison to other similar methods still requires due attention, although important results have been obtained in this area. Particularly interesting seems to be a study of the relationship between neural network and rough set approach to feature extraction from data.

Last but not least, rough set computer is badly needed for more serious computations in decision support. Some research in this area is already in progress.

For basic ideas of rough set theory the reader is referred to [11,15,19,20].

References

[1] E. Czogała, A. Mrózek, and Z. Pawlak. The idea of rough-fuzzy controller. *International Journal of Fuzzy Sets and Systems*, 72:61–63, 1995.

[2] D. Dubois and H. Prade. Putting rough sets and fuzzy sets together. In R. Słowiński, editor, *Intelligent Decision Support – Handbook of Advances and Applications of the Rough Set Theory*, pages 203–232. Kluwer Academic Publishers, Dordrecht, Boston, London, 1992.

[3] J. Grzymała-Busse. Knowledge acquisition under uncertainty - a rough set approach. *Journal of Intelligent and Robotics Systems*, 1:3–16, 1988.

[4] J. Grzymała-Busse. *Managing Uncertainty in Expert System.* Kluwer Academic Publishers, Dordrecht, Boston, London, 1991.

[5] J. Grzymała-Busse. Rough sets. *Advances in Imaging and Electrons Physics*, 94:to appear, 1995.

[6] E. Krusińska, R. Słowiński, and J. Stefanowski. Discriminant versus rough set approach to vague data analysis. *Journal of Applied Statistics and Data Analysis*, 8:43–56, 1992.

[7] T. Lin, editor. *CSC'95, 23rd Annual Computer Science Conference on Rough Sets and Database Mining*, San Jose, California, USA, March 1995. San Jose State University.

[8] T. Munakata. Rough control: A perspective. In *CSC'95, 23rd Annual Computer Science Conference on Rough Sets and Database Mining*, San Jose, California, USA, March 1995. San Jose State University.

[9] Z. Pawlak. Rough sets. *International Journal of Computer and Information Sciences*, 11:341–356, 1982.

[10] Z. Pawlak. *Rough Sets, Theoretical Aspects of Reasoning about Data.* Kluwer Academic Publishers, Dordrecht, Boston, London, 1991.

[11] Z. Pawlak, J. Grzymała-Busse, R. Słowiński, and W. Ziarko. Rough sets. *Communication of the ACM*, 38:88–95, 1995.

[12] Z. Pawlak and A. Skowron. Rough membership functions. In R. Yeager, M. Fedrizzi, and J. Kacprzyk, editors, *Advances in the Dempster-Shafer Theory of Evidence*, pages 251–271. John Wiley and Sons, New York, 1994.

[13] Z. Pawlak and R. Słowiński. Decision analysis using rough sets. *International Transactions on Operational Research*, 1:107–104, 1994.

[14] Z. Pawlak, S. Wong, and W. Ziarko. Rough sets: Probabilistic versus deterministic approach. *International Journal of Man Machine Studies*, 29:81–85, 1988.

[15] San Jose State University. *Proceedings of the Third International Workshop on Rough Sets and Soft Computing (RSSC'94)*, San Jose, California, USA, November 1994.

[16] A. Skowron and J. Grzymała-Busse. From rough set theory to evidence theory. In R. Yeager, M. Fedrizzi, and J. Kacprzyk, editors, *Advances in the Dempster-Shafer Theory of Evidence*, pages 193–236. John Wiley and Sons, New York, 1994.

[17] A. Skowron and C. Rauszer. The discernibility matrices and functions in information systems. In R. Słowiński, editor, *Intelligent Decision Support - Handbook of Advances and Applications of the Rough Set Theory*, pages 311–369. Kluwer Academic Publishers, Dordrecht, Boston, London, 1992.

[18] R. Słowiński, editor. *Intelligent Decision Support – Handbook of Advances and Applications of the Rough Set Theory.* Kluwer Academic Publishers, Dordrecht, Boston, London, 1992.

[19] R. Słowiński. Rough set approach to decision analysis. *AI Expert*, 10:18–25, 1995.

[20] A. Szladow and W. Ziarko. Rough sets: Working with imperfect data. *AI Expert*, 8:36–41, 1993.

[21] W. Ziarko, editor. *Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceeding of the International Workshop on Rough sets and Knowledge Discovery (RSKD'93)*, Banff, Alberta, Canada, 1993. Springer Verlag, Berlin, Heidelberg, New York, London , Paris, Tokyo, Hong Kong, Barcelona, Budapest.

[22] W. Ziarko. Variable precision rough set model. *Journal of Computer and System Sciences*, 40:39–59, 1993.