

ROUGH SET APPROACH TO KNOWLEDGE-BASED DECISION SUPPORT

Zdzislaw PAWLAK *

Abstract. The rough set concept is a new mathematical approach to imprecision, vagueness and uncertainty. To some extent it overlaps 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 presents the basic assumptions underlying the rough sets philosophy, gives its fundamental concepts and discusses briefly some areas of applications, in particular in decision support. Finally further problems are shortly outlined.

1 Introduction

1.1 Basic Philosophy

The rough set concept proposed by the author in [5] is a new mathematical approach to imprecision, vagueness and uncertainty. The rough set philosophy is founded on the assumption that with every objects 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. Objects characterized by the same information are indiscernible (similar) in view of the available information 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 as members of the set or 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.

*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 Sciences, ul. Baltycka 5, 44 000 Gliwice, Poland

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 concepts, cannot be characterized in terms of information about their elements. Therefore 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.

1.2 The Relationship to Other Theories

The rough set concept overlaps – to some extent – with many other mathematical tools developed to deal with imperfect knowledge.

Frequently the rough set theory is contrasted with the fuzzy set theory. Basically the idea of fuzzy set and rough set are not competitive, but complementary since they refer to different aspects of imprecision, and consequently are meant to be used in different areas. In the fuzzy set theory imprecision is expressed by a membership function, whereas the rough set approach is based on indiscernibility and approximations [7].

Another relationship exists between the rough set theory and Dempster-Shafer theory of evidence [10]. The main difference is that Dempster-Shafer theory uses belief functions as a main tool, while rough set theory makes use of sets – lower and upper approximations.

Furthermore, some relations exist between the rough set theory and discriminant analysis [3] and the Boolean reasoning methods [11].

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

1.3 Some Remarks on Applications

The 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 and pattern recognition. It seems of particular importance to 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, or basic probability assignment in Dempster-Shafer theory, grade of membership or the value of possibility in fuzzy set theory. An extensive study of various mathematical models of uncertainty can be found in [1].

For basic ideas of the rough set theory the reader is referred to [6]. Many interesting applications of this approach are presented and discussed in [4,12,14]. The relationship between the rough set theory and decision analysis is presented in [8]. The present state of the rough set theory and its further perspectives are discussed in [9].

2 Basic Concepts of the Rough Set Theory

2.1 Indiscernibility Relation

As we mentioned in the introduction basic operations in the rough set theory are approximations of sets. In this section we define these concepts more precisely.

Suppose we are given a finite set of objects U called the *universe* and a binary relation I over U called the *indiscernibility* relation. The indiscernibility relation can be an equivalence or the tolerance relation. The second is more general than the first one but for the sake of simplicity we assume that I is an equivalence relation. The relation I is meant to express the fact that our knowledge about elements of the universe is limited and therefore we are unable, in general, to discern them. That means that, in general, we are unable to deal with single elements of the universe but with clusters of similar elements (atoms), which from the mathematical point of view are equivalence classes of the indiscernibility relation I .

2.2 Approximations

Let $X \subseteq U$. The *I-lower* and the *I-upper approximation* of X are defined thus:

$$\begin{aligned} I_*(X) &= \{x \in U : I(x) \subseteq X\}, \\ I^*(X) &= \{x \in U : I(x) \cap X \neq \emptyset\}, \end{aligned}$$

where $I(x)$ denotes the set of all objects indiscernible with x , i.e. equivalence class determined by 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 .

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 . Thus the coefficient $\alpha_I(X)$ can be understood as the accuracy of approximation of the concept X .

2.3 Rough Membership

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.

2.4 Reduction of Knowledge and Dependencies

Usually we need many classification patterns of objects. For example patients can be classified according to temperature, blood pressure etc. Hence we can assume that we have not one, but a family of indiscernibility relations $\mathbf{I} = \{I_1, I_2, \dots, I_n\}$ over the universe U . Set theoretical intersection of equivalence relations I_1, I_2, \dots, I_n , denoted

$$\bigcap \mathbf{I} = \bigcap_{i=1}^n I_i,$$

is also an equivalence relation. In this case elementary sets are equivalence classes of the equivalence relation $\bigcap \mathbf{I}$. Because elementary sets uniquely determine our knowledge about the universe, the question arises whether some classification patterns can be removed without changing the family of elementary sets - or in other words, preserving the indiscernibility. Minimal subset \mathbf{I}' of \mathbf{I} such that $\bigcap \mathbf{I} = \bigcap \mathbf{I}'$ will be called a *reduct* of \mathbf{I} .

Of course \mathbf{I} can have many reducts. Finding reducts is not a very simple task and there are many methods to solve this problem, however we will not consider them

here, but explain it by means of example in the next section. Readers interested in precise mathematical formulation of this problem are referred to [6] and [11].

Another important issue is the relationship between various classification patterns, i.e. between equivalence relations in the family \mathbf{I} . We will say that indiscernibility relation I depends on the indiscernibility relation I' , $I' \rightarrow I$, if $I' \subseteq I$, i.e. each equivalence class of I' is included in some equivalence class of I . In other words it means that if $I' \rightarrow I$, then every elementary set generated by I' is included in some elementary set generated by I . That is to say that any granule of knowledge associated with I' is a part of some granule of knowledge associated with I . Thus dependency explains relationship between various classification patterns.

3 Illustrative Example

Departure points of the rough set philosophy from the computational point of view are data tables, called also attribute-value tables or information systems, or decision tables, when condition and decision attributes are distinguished. Basic problems which can be solved using the rough set approach are the following:

- 1) description of objects in terms of attribute values
- 2) dependencies (full or partial) between attributes
- 3) reduction of attributes
- 4) significance of attributes
- 5) decision rules generation

and others.

All the above mentioned problems can be formulated in terms of the two basis operations in the rough set theory, the lower and the upper approximations, which will be illustrated by the following example. Suppose we are given data about 6 patients, as shown in Table 1.

Patient	Headache	Muscle-pain	Temperature	Flu
p1	no	yes	high	yes
p2	yes	no	high	yes
p3	yes	yes	very high	yes
p4	no	yes	normal	no
p5	yes	no	high	no
p6	no	yes	very high	yes

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 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 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 p2 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 can not 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.

The accuracy coefficient for the concept "flu" is

$$\alpha_I(flu) = \frac{|\{p1, p3, p6\}|}{|\{p1, p2, p3, p5, p6\}|} = 3/5,$$

and for the concept "not flu" we have

$$\alpha_I(not\,flu) = \frac{|\{p4\}|}{|\{p2, p4, p5\}|} = 1/3.$$

We can also compute membership value for each patient to the concept "flu" or "not flu". The numerical values of the membership function are given below:

$$\begin{aligned} \mu_{flu}^I(p1) &= \frac{|\{p1, p2, p3, p6\} \cap \{p1\}|}{|\{p1\}|} = 1, \\ \mu_{flu}^I(p2) &= \frac{|\{p1, p2, p3, p6\} \cap \{p2, p4\}|}{|\{p2, p4\}|} = 1/2, \\ \mu_{flu}^I(p3) &= \frac{|\{p1, p2, p3, p6\} \cap \{p3\}|}{|\{p3\}|} = 1, \end{aligned}$$

$$\begin{aligned}\mu_{flu}^I(p4) &= \frac{|\{p1, p2, p3, p6\} \cap \{p4\}|}{|\{p4\}|} = 0, \\ \mu_{flu}^I(p5) &= \frac{|\{p1, p2, p3, p6\} \cap \{p2, p5\}|}{|\{p2, p5\}|} = 1/2, \\ \mu_{flu}^I(p6) &= \frac{|\{p1, p2, p3, p6\} \cap \{p6\}|}{|\{p6\}|} = 1.\end{aligned}$$

The interpretation of the accuracy coefficients and the membership grades is left to the interested reader.

We may also ask whether all attributes in this table are necessary to define flu. One can easily see, for example that, if a patient has very high temperature, he has for sure flu, but if he has normal temperature he has not flu whatsoever.

The problem of elimination of superfluous attributes boils down to finding so called reducts of the whole set of attributes.

One can compute that for the example shown in Tab.1 we have two reducts: {Headache, Temperature} and {Muscle-pain, Temperature}. That means that either the attribute Headache or Muscle-pain can be eliminated from the table without changing its elementary sets. Hence instead of Tab 1. we can use either Tab.2

Patient	Headache	Temperature	Flu
p1	no	high	yes
p2	yes	high	yes
p3	yes	very high	yes
p4	no	normal	no
p5	yes	high	no
p6	no	very high	yes

Tab.2

or Tab.3

Patient	Muscle-pain	Temperature	Flu
p1	yes	high	yes
p2	no	high	yes
p3	yes	very high	yes
p4	yes	normal	no
p5	no	high	no
p6	yes	very high	yes

Tab.3

Either table can be used equivalently to analyze the decisions without losing any information in comparison to Tab.1.

The tables considered above are also known as decision tables, where Headache, Muscle-pain and Temperature are referred to as condition attributes whereas Flu is called the decision attribute. Each decision table can be represented in the form

of a decision algorithm, consisting of decision rules of the form "IF ... THEN ...". For example Tab.2 can be represented in the form:

IF (Headache,no) & (Temperature, high) → THEN (Flu,yes)
 IF (Headache,yes) & (Temperature, high) → THEN (Flu,yes)
 IF (Headache,yes) & (Temperature, very high) → THEN (Flu,yes)
 IF (Headache,no) & (Temperature, normal) → THEN (Flu,no)
 IF (Headache,yes) & (Temperature, high) → THEN (Flu,no)
 IF (Headache,no) & (Temperature, very high) → THEN (Flu,yes).

Using the rough set technique (cf. [6]) the decision algorithm can be simplified further as shown below:

IF (Headache,no) & (Temperature, high) → THEN(Flu,yes)
 IF (Headache,yes) & (Temperature, high) → THEN (Flu,yes)
 IF (Temperature, very high) → THEN (Flu,yes)
 IF (Temperature, normal) → THEN (Flu,no)
 IF (Headache,yes) & (Temperature, high) → THEN (Flu,no).

Let us observe that rules

IF (Headache,yes) & (Temperature, high) → THEN (Flu,yes)
 IF (Headache,yes) & (Temperature, high) → THEN (Flu,no)

have the same conditions but different decisions, so that proper decision cannot be made by applying this kind of rules. They are called inconsistent (nondeterministic, conflicting) rules. The remaining three rules are consistent (deterministic). Consistent and inconsistent rules are also often called sure (certain) and possible rules, respectively. Sure decision rules determine uniquely decisions, whereas possible decision rules determine set of possible decision, which can be made on the basis of given conditions. With every possible decision rule we will associate a credibility factor, which evaluates the credibility of each possible decision suggested by the rule. There are many ways to define this coefficient. We propose here to employ the membership function to this end. Any decision rule is generated from a row of the decision table. It turns out that the membership function value for the object associated with the decision rule can be used as a basis to define the credibility of decision rules.

Let $\delta(x)$ denote the decision rule associated with object x . We will say also that x supports rule $\delta(x)$. Then the credibility factor of this rule can be defined as

$$C(\delta(x)) = \begin{cases} 1, & \text{if } \mu_X^I(x) = 0 \text{ or } 1 \\ \mu_X^I(x), & \text{if } 0 < \mu_X^I(x) < 1. \end{cases}$$

In this way any consistent rule will get one as a credibility factor value, while any inconsistent rule will get credibility factor, which is smaller than one (but not equal to zero). The closer the credibility factor is to one the more credible is the rule.

If a decision rule is supported by more than one object, then as the credibility factor coefficient we assume the credibility factor of the most supporting object. The credibility factor for both inconsistent rules in the example above is 0.5. This coefficient can be seen as an alternative to Dempster-Shafer approach [10], where belief and plausibility functions are used.

In fact the above considerations give rise to the question whether Flu depends on symptoms Headache, Muscle-pain and Temperature. As the above analysis shows this is not the case in our example, since the inconsistent rules do not allow to give definite answer. In order to tackle with this kind of situations partial dependency of attributes can be defined [6]. Partial dependency is expressed by the degree to which the dependency between attributes hold and is defined simple as the ratio of the number of deterministic rules to the number of all rules in the decision table. In our example the ratio is $4/6 = 2/3$. If all decision rules were deterministic then the dependency coefficient would be one and the dependency is called in this case total.

4 Rough Sets and Decision Analysis

Any decision problem involves a set of objects, e.g. actions, states, processes, competitors, etc. In general objects can be anything we can think of. The objects are described by a attribute-value pairs. As we mentioned already, such sets of data can be represented by a table rows of which correspond to objects, columns to attributes and entries of the table are attribute value.

The data table represents some facts about the decision problem. In particular it may represent opinions of agents, groups of agents, decision makers, etc. In the table two classes of attributes are usually distinguished – called usually condition and decision attributes. Such tables are known as decision tables. Each row of the decision table specifies a decision rule, which determines decision (action) which should be performed if specific conditions are satisfied. In other words the decision table can be viewed as a collection of "if...than..." decision rules.

The aim of the decision analysis is to answer the following two basic questions. The first: question is to *explain* decision in terms of circumstances in which the decision has been made. The second is to give *prescription* how to make decision under specific circumstances. Prescription is mainly based on decision rules derived from decision table [2,13]. In this sens, the rough set approach is similar to inductive learning approach, however, the former one is going far beyond the latter because in the rough set approach, the prescription task is preceded by the explanation which gives pertinent information useful decision support. Besides, optimization of decision rules as well discovering dependencies between rules, and elimination of superfluous ones is also of great importance.

The rough set theory offers mathematical tools to answer the above mentioned questions, and seems particularly suited to analyze this kind of problems. In other words the rough set theory offers technic to generate optimal set of decision rules from specification of the decision process.

More about the application of the rough set theory in decision analysis can be found in [8].

5 Conclusion

Many software systems for decision support based on the rough set theory have been implemented in many countries, and successfully applied to many real-life problems in various areas, e.g. medicine, pharmacology, business, banking, market research, engineering design, meteorology, vibration analysis, conflict analysis, image processing, voice recognition, concurrent system analysis and others and others. Some of them are available commercially.

The rough set theory turned out to be a very useful tool for decision support systems, especially when vague concepts and uncertain data are involved in the decision process. Nevertheless several theoretical and practical problems require further attention. The most important one seems to develop a decision logic for imprecise reasoning, based on the rough set philosophy. A great deal of effort has been devoted to pursue this problem, but it remains still open.

Another important issue is efficient software development for rough set based data analysis. Present software products are not always satisfactory, particularly if large collections of data are to be analyzed.

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.

References

- [1] Grzymala-Busse J.W.: *Managing Uncertainty in Expert Systems*. Kluwer Academic Publishers, Dordrecht, Boston, London, 1991.
- [2] Grzymala-Busse J. W. : LERS-a system for learning from examples based on rough sets. In: R.Slowinski (ed.), *Intelligent Decision Support - Handbook of Applications and Advances in the Rough Set Theory*. Kluwer Academic Publishers, Dordrecht, 1992, pp. 3-18.
- [3] Krusinska E., Slowinski R., Stefanowski J.: Discriminant Versus Rough Set Approach to Vague Data Analysis. *Applied Stochastic Models and Data Analysis* **8** (1992) 43- 56.
- [4] Lin T.Y. ed.: *The Third International Workshop on Rough Sets and Soft Computing Proceedings (RSSC'94)*. San Jose State University, San Jose, California, USA, November 10-12, 1994.
- [5] Pawlak Z.: *Rough sets*. *International Journal of Computer and Information Sciences* **11**, (1982) 341-356.
- [6] Pawlak Z.: *Rough Sets - Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht, Boston, London, 1991.

- [7] Pawlak Z., Skowron A.: Rough Membership Functions. In: Yaeger R.R., Fedrizzi M. and Kacprzyk J., eds. *Advances in the Dempster Shafer Theory of Evidence*, John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore (1994), 251-271.
- [8] Pawlak Z., Slowinski R.: Rough Set Approach to Multi-Attribute Decision Analysis, Invited Review. *European Journal of Operational Research* **72** (1994) 443-459.
- [9] Pawlak Z., Grzymala-Busse J. W., Slowinski R., Ziarko, W.: Rough Sets, *Communication of the ACM* (to appear).
- [10] Skowron A., Grzymala-Busse J.W.: From 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 & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore (1994), 193-236.
- [11] Skowron A., Rauszer, C.: 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, Dordrecht, Boston, London (1992), 311-362.
- [12] Slowinski R. ed.: *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*. Kluwer Academic Publishers, Dordrecht, Boston, London, 1992.
- [13] Slowinski R., Stefanowski J.: Rough classification with valued closeness relation. In: E.Diday, Y.Lechevallier, M.Schrader, P.Bertrand, B.Burtschy (eds.), *New Approaches in Classification and Data Analysis*. Springer-Verlag, Berlin, 1994, pp. 482-489.
- [14] Ziarko W. ed.: *Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery (RSKD'93)*. Banff, Alberta, Canada, 12-15 October 1993, Springer Verlag, 1993.