ROUGH SET THEORY AND ITS APPLICATIONS TO DATA ANALYSIS

Zdzisław Pawlak

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

ABSTRACT

The paper gives basic ideas of rough set theory - a new approach to data analysis. The lower and the upper approximation of a set the basic operations of the theory, are intuitively explained and formally defined. Some applications of rough set theory are briefly outline and some future problems pointed out.

INTRODUCTION

Rough set theory [7] is a new mathematical approach to data analysis and data mining. After 15 year of pursuing of rough set theory and its application the theory has reached a certain degree of maturity. In recent years we witnessed a rapid grow of interest in rough set theory and its application, world wide. Many international workshops, conferences and seminars included rough sets in their programs. A large number of high quality papers have been published recently on various aspects of rough sets.

The notion of a set is a fundamental concept for a whole contemporary mathematics, and set theory introduced by George Cantor in 1883 is, no doubt, a mile stone in development of modern mathematical thinking.

A set is defined by its elements, i.e. it is defined if all its elements are uniquely determined. For example, the set of all odd (even) numbers is determined uniquely and every integer can be classified, without any doubt, as „odd" or „even". This kind of notions are usually refereed to as *crisp*. Obviously all mathematical notions are crisp, otherwise it would be impossible to prove any mathematical theorem.

But in many other fields the situation in not so pleasant. For example, in medicine the notion of a "healthy (or ill) person" cannot be uniquely defined. Similarly in law the notion of „guilty" („not guilty") cannot be precisely defined. These kind of imprecise notions are called *vague*. Vague notions are notoriously used, not only in medicine or law, but almost everywhere; for example, economy and politics are other exemplary domains where vague concepts are intrinsically adhered to method of thinking and debates.

Vague concepts are characterized by a „boundary region", which consists of all elements which cannot be classified to the concept or its complement. For example, the concept of an odd (even) number is precise, because every number is either odd or even - whereas the concept of a beautiful women is vague, because for some women we cannot decide, with certainty, whether they are beautiful or not.

This approach is know in a philosophical literature as boundary-line approach to vagueness and is attributed to German logician Gotlob Frege, who first formulate this idea in 1894.

Thus vague concepts form basis for *common sense reasoning* in many fields connected with real life situations.

Vagueness for many years attracted attention of philosophers and logicians. Recently, computer scientist also got interested in vagueness, for many computer applications, in particular referring to „artificial intelligence", badly need use of vague notions, and vague concepts based reasoning methods.

The most successful theoretical approach to vagueness is no doubt fuzzy set theory proposed by Zadeh. Basic idea of fuzzy set theory hinges on fuzzy membership function, which allows partial membership of elements to a set, i.e. it allows that elements can belong to a set to „a degree".

Rough set theory is another mathematical approach to vagueness.

In the rough set approach vagueness is due to the lack of information, about some elements of the universe.

If with some elements the same information is associated, in view of this information these elements are *indiscernible*. For example, if some patients suffering from a certain disease display the same symptoms, they are indiscernible with respect to the information about them. It turns out that the indiscernibility leads to the boundary-line cases, i.e. that some elements cannot be classified to the concept or its complement in view of the available information.

Because vague concepts have boundary-line cases, i.e. elements which cannot be with certainty classified as elements of the concept, hence vagueness is strictly connected with the idea of *certainty* (or *uncertainty*).

Rough set theory seems to be well suited as a mathematical model of vagueness and uncertainty. Vagueness is a property of sets (concepts) and is strictly related to the existence of the boundary region of a set, whereas uncertainty is a property of elements of sets. In the rough set approach both concepts are closely related and are due to the indiscernibility caused by insufficient information about the world we are interested in.

The connection of rough set theory and many other theories has been clarified. Particularly interesting is the relationship between 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 [10], whereas the connection with theory of evidence is more substantial [12]. Besides, rough set theory is related to discriminate analysis [4], Boolean reasoning methods [13] and others. The relationship between rough set theory and decision analysis is presented in [11,15]. Several extension of the "basic" model of rough set have been proposed and investigated.

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 and data mining. 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. Rough set theory has been successfully applied in many real-life problems in medicine, pharmacology, engineering, banking, financial and market analysis and others. In particular, in pharmacology the analysis of relationships between the chemical structure and the antimicrobial activity of drugs has been successfully investigated. Banking applications include evaluation of a bankruptcy risk 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 material sciences, diagnosis of machines using vibroacoustics symptoms (noise, vibrations) and process control. Application in linguistics, environment and databases are other important domains. More about applications of rough set theory can be found in [5, 6,9,14,17,18,20,22] and others.

Very promising new areas of application of the rough set concept seems to emerge in the near future. They include rough control, rough data bases, rough information retrieval, rough neural network and others.

Application of rough sets requires a suitable software. Many software systems for workstations and personal computers based on rough set theory have been developed. The most known include LERS [1], Rough DAS and Rough Class [16] and DATALOGIC [17]. Some of them are available commercially.

BASIC PHILOSOPHY

Rough set philosophy is based on the assumption that, in contrast to the classical set theory, we have some additional information (knowledge, data) about elements of a set. Consider, for example, a group of patients suffering from a certain disease. In a hospital treating the patients there are data files containing information about patients – such as, e.g., body temperature, blood pressure, name, age, address and others. 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 (atoms) of our knowledge about reality we are interested in. 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*). With every set X we can associate two crisp sets, called the *lower* and the *upper approximation* of X. The lower approximation of X is the union of all elementary set which are included in X, whereas the upper approximation of X is the union of all elementary set which have non-empty intersection with X. In other words the lower approximation of a set is the set of all elements that *surely* belongs to X, whereas the upper approximation of X is the set of all elements that *possibly* belong to X. The difference of the upper and the lower approximation of X is its *boundary region*. Obviously a set is rough if it has non empty boundary region whatsoever; otherwise the set is crisp. Elements of the boundary region cannot be classified, employing the available knowledge, either to the set or its complement. Approximations of sets are basic operation in rough set theory and are used as main tools to deal with vague and uncertain data.

EXAMPLE

Let us depict the above idea by means of a simple example.

Data are often presented as a table, columns of which are labeled by *attributes*, rows by *objects* of interest and entries of the table are *attribute values*. Such tables are known as *information systems, attribute-value tables*, *data tables* or *information tables*.

Usually we distinguish in information tables two kinds of attributes, called *condition* and *decision* attributes. Such tables are known as *decision tables.* Rows of a decision table are referred to as "if...then..." *decision rules*, which give conditions necessary to make decisions specified by the decision attributes. An example of a decision table is shown in Table 1.

Pipe	*C*	*S*	*P*	*Cracks*
1	*high*	*high*	*low*	*yes*
2	*avg.*	*high*	*low*	*no*
3	*avg.*	*high*	*low*	*yes*
4	*low*	*low*	*low*	*no*
5	*avg.*	*low*	*high*	*no*
6	*high*	*low*	*high*	*yes*

Table 1

The table contains data concerning six cast iron pipes exposed to high pressure endurance test. In the table *C*, *S* and *P* are condition attributes, displaying the percentage content in the pig-iron of coal, sulfur and phosphorus respectively, whereas the attribute *Cracks* revels the result of the test. The values of condition attributes are as follows (*C*, high) > 3.6%, 3.5% ≤ (*C*, avg.) ≤ 3.6%, (*C*, low) < 3.5%, (*S*, high) ≥ 0.1%, (*S*, low) < 0.1%, (*P*, high) ≥ 0.3%, (*P*, low) < 0.3%.

Main problem we are interested in is how the endurance of the pipes depend on the compounds *C*, *S* and *P* comprised in the pig-iron, or in other words, if there is a functional dependency between the decision attribute *Cracks* and the condition attributes *C, S* and *P.* In rough set theory language this boils down to the question, if the set {2,4,5}of all pipes having no cracks after the test (or the set {1,3,6}of pipes having cracks), can be uniquely defined in terms of condition attributes values.

It can be easily seen that this is impossible, since pipes 2 and 3 display the same features in terms of attributes *C*, *S* and *P*, but they have different values of the attribute *Cracks*. Thus information given in Table 1 is not sufficient to solve our problem. However we can give a partial solution. Let us observe that if the attribute *C* has the value *high* for a certain pipe, then the pipe have cracks, whereas if the value of the attribute *C* is *low*, then the pipe has no cracks. Hence employing attributes *C*, *S* and *P*, we can say that pipes 1 and 6 are *surly* good, i.e., *surely* belong to the set {1, 3, 6}, whereas pipes 1, 2, 3 and 6 are *possible* good, i.e., possible belong to the set {1, 3, 6}.Thus the sets {1, 6}, {1, 2, 3, 6} and {2, 3} are the lower, the upper approximation and the boundary region of the set {1, 3, 6} respectively.

This means that the quality of pipes cannot be determined exactly by the content of coal, sulfur and phosphorus in the pig-iron, but can be determined only with some approximation.

In fact approximations determine the dependency (total or partial) between condition and decision attributes, i.e., express relationship between values of condition and decision attributes. The degree of dependency between condition and decision attributes can be informally defined as the ratio of all rows in which values of condition attributes uniquely determine values of decision attributes – to all rows in the table. For example, the degree of dependency between cracks and the composition of the pig-iron is 4/6 = 2/3. That means that four out of six (ca. 60%) pipes can be properly classified as good on the basis of their composition.

We might be also interested in reducing some of the condition attributes, i.e. to know whether all conditions are necessary to make decisions specified in a table. To this end we will employ the notion of a *reduct* (of condition attributes). By a reduct we understand a minimal subset of condition attributes which preserves degree of dependency between decision and condition attributes. It is easy to compute that in Table 1 we have two reducts {*C, S*} and {*C, P*}. Intersection of all reducts is called the *core.* In our example the core is the attribute *C*.

That means that in view of the data coal is the most important factor causing cracks and cannot be eliminated from our considerations, whereas sulfur and phosphorus play a minor role and can be mutually exchanged as factors causing cracks.

INDISCERNIBILITY

As mentioned in the introduction, the starting point of rough set theory is the indiscernibility relation, generated by information about objects of interest. The indiscernibility relation is intended to express the fact that due to the lack of knowledge we are unable to discern some objects employing the available information. That means that, in general, we are unable to deal with single objects but we have to consider clusters of indiscernible objects, as fundamental concepts of knowledge.

Now we present above considerations 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*.

The pair $S = (U, A)$ will be called an *information system*. 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_By 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., the partition determined by B, will be denoted by U/I_B, or simply U/B; an equivalence class of I_B, i.e., the block of the 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 referred to as *B-elementary concepts* or *B-granules*. As mentioned previously in the rough set approach the elementary concepts are the basic building blocks (concepts) of our knowledge about reality.

For example pipes 1, 2 and 3, as well as pipes 5 and 6, are indiscernible in terms of S and P; pipes 1 and 2, as well 2, 3 and 5 are indiscernible in terms of attribute C.

APPROXIMATIONS OF SETS

The indiscernibility relation will be used next to define basic concepts of rough set theory. Let us define now the following two operations on sets

$$B_*(X) = \{x \in U : B(x) \subseteq X\},$$
$$B^*(X) = \{x \in U : B(x) \cap X \neq \varnothing\},$$

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

The example which follows will show how to compute the lower, the upper and the boundary region of a set.

Let us denote the set of an condition attributes C, S and P by B, and let $X_{yes} = \{1, 3, 6\}$, $X_{no} = \{2, 4, 5\}$ denote sets of pipes having cracks and having no cracks respectively. B-elementary sets in our example are the following sets: $\{1\}$, $\{2, 3\}$, $\{4\}$, $\{5\}$ and $\{6\}$.

Hence we get

$$B_*(X_{yes}) = \{1\} \cup \{6\} = \{1,6\},$$
$$B^*(X_{yes}) = \{1\} \cup \{2,3\} \cup \{2,3\} \cup \{6\} = \{1,2,3,6\},$$
$$B_*(X_{no}) = \{4\} \cup \{5\} = \{4,5\},$$
$$B^*(X_{no} = \{2,3\} \cup \{4\} \cup \{5\} = \{2,3,4,5\},$$
$$BN_B(X_{yes}) = BN_B(X_{no}) = \{2,3\}.$$

One can easily show the following properties of approximations:

(1) $B_*(X) \subseteq X \subseteq B^*(X)$,

(2) $B_*(\varnothing) = B^*(\varnothing) = \varnothing, B_*(U) = B^*(U) = U$,

(3) $B^*(X \cup Y) = B^*(X) \cup B^*(Y)$,

(4) $B_*(X \cap Y) = B_*(X \cap B_*(Y)$,

(5) $X \subseteq Y$ implies $B_*(X) \subseteq B_*(\mathrm{Y})$ and $B^*(X) \subseteq B^*(Y)$,

(6) $B_*(X \cup Y) \supseteq B_*(X) \cup B_*(Y)$,

(7) $B^*(X \cap Y) \subseteq B^*(X) \cap B^*(Y)$,

(8) $B_*(-X) = -B^*(X)$,

(9) $B^*(-X) = -B_*(X)$,

(10) $B_*(B_*(X)) = B^*(B_*(X)) = B_*(X)$,

(11) $B^*(B^*(X)) = B_*(B^*(X)) = B^*(X)$,

where $-X$ denotes $U - X$.

Let us observe that due to the properties (6) and (7) approximations cannot be computed step by step, because the lower approximation of union of sets is not necessarily equal to union of the lower approximations of the constituent sets; similarly the upper approximation of intersection of sets needs to be equal to the intersection of the upper approximation of the constituent sets.
That means, that in general, the data table cannot be decomposed into smaller parts (or that tables cannot be combined together), for the results obtained from decomposed (or integrated) tables may differ.

One can define the following four basic classes of rough sets, i.e., four categories of vagueness:

a) $B_*(X) \neq \varnothing$ and $B^*(X) \neq U$, iff X is *roughly B-definable*,

b) $B_*(X) = \varnothing$ and $B^*(X) \neq U$, iff X is *internally B-indefinable*,

c) $B_*(X) \neq \varnothing$ and $B_*(X) = U$, iff X is *externally B-definable*,

d) $B_*(X) = \varnothing$ and $B^*(X) = U$, iff X is *totally B-indefinable*.

The intuitive meaning of this classification is the following.

If X is roughly *B-definable*, this means that we are able to decide for some elements of U whether they belong to X or $-X$, using B.

If X is internally B-indefinable, this means that we are able to decide whether some elements of U belong to $-X$, but we are unable to decide for any element of U, whether it belongs to X or not, using B.

If X is externally B-indefinable, this means that we are able to decide for some elements of U whether they belong to X, but we are unable to decide, for any element of U whether it belongs to $-X$ or not, using B.

If X is totally B-indefinable, we are unable to decide for any element of U whether it belongs to X or $-X$, using B.

In our example $B_*(X_{yes}) \neq U$, $B^*(X_{yes}) \neq \varnothing$ and $B_*(X_{no}) \neq U$, $B^*(X_{no}) \neq \varnothing$ hence both X_{yes} and X_{no} are roughly B-definable.

Rough set can be also characterized numerically by the following coefficient

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

called the *accuracy* of *approximation*, where $|X|$ denotes the cardinality of X. Obviously $0 \le \alpha_B(X) \le 1$. 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.

For example, for X_{yes} we have

$$\alpha_B(X_{yes}) = \alpha_B(X_{no}) = 1/4 = 1/2.$$

ROUGH SETS AND MEMBERSHIP FUNCTION

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*).

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

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

The rough membership function has the following properties [35]:

a) $\mu_X^B(x) = 1$ iff $x \in B_*(X)$,

b) $\mu_X^B(x) = 0$ iff $x \in -B^*(X)$,

c) $0 < \mu_X^B(x) < 1$ iff $x \in BN_B(X)$,

d) If $(B) = \{(x,x) : x \in U\}$, then $\mu_X^B(x)$ is the characteristic function of X,

e) If $xI(B)y$, then $\mu_X^B(x) = \mu_X^B(y)$ provided $I(B)$,

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

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

h) $\mu_{X \cap Y}^B(x) \le \min(\mu_Y^B(x), \mu_Y^B(x))$ for any $x \in U$,

i) If **X** is a family of pairwise disjoint sets of U, then $\mu_{\cup X}^B(x) = \sum_{X \in \mathrm{X}} \mu_X^B(x)$ for any $x \in U$.

The above properties show clearly the difference between fuzzy and rough memberships. In particular properties g) and h) show that the rough membership can be regarded formally as a generalization of fuzzy membership, for the max and the min operations for union and intersection of sets respectively for fuzzy sets are special cases of that for rough sets. But let us recall that the „rough membership", in contrast to the „fuzzy membership", has probabilistic flavor.

For example the rough membership of pipes to X_{yes} is shown below:

$$\mu^B_{X_{yes}}(p_1) = 1, \qquad \mu^B_{X_{yes}}(p_4) = 0,$$
$$\mu^B_{X_{yes}}(p_2) = 1/2, \qquad \mu^B_{X_{yes}}(p_5) = 0,$$
$$\mu^B_{X_{yes}}(p_3) = 1/2, \qquad \mu^B_{X_{yes}}(p_6) = 1.$$

That means that if, for example, a pipe has high content of coal and sulfur and low content of phosphorus, then it has cracks, whereas if content of coal, sulfur and phosphorus is average, high and low respectively, the probability that the pipe has cracks equals 1/2.

It can be easily seen that there exists a strict connection between vagueness and uncertainty. As we mentioned above vagueness is related to sets (concepts), whereas uncertainty is related to elements of sets.

DEPENDENCY OF ATTRIBUTES

Approximations of sets are strictly related with the concept of dependency (total or partial) of 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 value of attribute form *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 no total dependencies whatsoever.

We would also need a more general concept of dependency of attributes, called the *partial dependency* of attributes. Partial dependency means that only some values of *D* are determined by values of *C*.

Formally dependency can be defined in the following way. Let *D* and *C* be subsets of *A*.

We will say that *D depends on C* in a *degree* $k(0 \le k \le 1)$, denoted $C \Rightarrow_k D$, if

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

where

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

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*.
Obviously

$$\gamma(C, D) = \sum_{X \in U/D} \frac{|C_*(X)|}{|U|}.$$

If $k = 1$ we say that *D depends totally* on *C*, and if $k < 1$, we say that *D depends partially* (in a *degree k*) on *C*.

The coefficient *k* expresses the ratio of all elements of the universe, which can be properly classified to block of the partition *U/D*, employing attributes *C* and will be called the *degree of the dependency* which can be also interpreted as a probability that $x \in U$ belongs to one of the decision classed determined by decision attributes

For example $\{Cracks\}$ depends on $\{C, S, P\}$ in the degree $k = 4/6 = 2/3$.

That means that only two third of pipes can be property classified as having cracks or not using attributes *C*, *S* and *P*.

REDUCTION OF ATTRIBUTES

We often face a question whether we can remove some data from a data-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 *S* or *P* we get the data set which is equivalent to the original one, with regard to lower approximations and the degree of the dependency.

Let us express this idea more precisely.

Let $C, D \subseteq A$, be sets of condition and decision attributes, respectively. We will say that $C' \subseteq C$ is a *D-reduct* (reduct with *respect* to *D*) of *C*, if C' is a minimal subset of *C* such that

$$\gamma(C, D) = \gamma(C', D).$$

For example in Table 1 we have two reducts $\{C, S\}$ and $\{C, P\}$with respect to attribute *Cracks*. That means, that either the attribute *S* or *P* can be eliminated from the table and consequently instead of Table 1 we can use either Table 2

Pipe	*C*	*S*	*Cracks*
1	*high*	*high*	*yes*
2	*avg.*	*high*	*no*
3	*avg.*	*high*	*yes*
4	*low*	*low*	*no*
5	*avg.*	*low*	*no*
6	*high*	*low*	*yes*

Table 2

or Table 3

Pipe	*C*	*P*	*Cracks*
1	*high*	*low*	*yes*
2	*avg.*	*low*	*no*
3	*avg.*	*low*	*yes*
4	*low*	*low*	*no*
5	*avg.*	*high*	*no*
6	*high*	*high*	*yes*

Table 3

The intersection of all *D*-reduct is called *D-core* (core with *respect* to *D*).
Because the core is the intersection of all reducts, it is included in every reduct, i.e., each element of the core belongs to some reduct. Thus, in a sense, the core is the most important subset of attributes, for none of its elements can be removed without affecting of the classification power of attributes. In Table 1 the core of $\{C, S, P\}$, with respect to *Cracks* is *C*.

DISCERNIBILITY MATRICES AND FUNCTIONS

To compute easily reducts and the core we will use discernibility matrix [42], which is defined next.

By a discernibility matrix of $B \subseteq A$ we will mean $n \times n$ matrix defined as:

$$\delta(x, y) = \{a \in B: a(x) \neq a(y)\}.$$

Thus $\delta(x,y)$ is the set of all attributes which discern objects x and y.

The discernibility matrix for Table 1 and the set of attributes $B = \{C, S, P\}$ is given in Table 4.

	1	2	3	4	5	6
1						
2	*C*					
3	*C*	∅				
4	*C*,*S*	*C*,*S*	*C*,*S*			
5	*C*,*S*,*P*	*S*,*P*	*S*,*P*	*C*,*P*		
6	*S*,*P*	*C*,*S*,*P*	*C*,*S*,*P*	*C*,*P*	*C*	

Table 4

The discernibility matrix assigns to each pair of objects x and y a subset of attributes $\delta(x, y) \subseteq B$, with the following properties:

i) $\delta(x, y) = \emptyset$,

ii) $\delta(x, y) = \delta(y, x)$,

iii) $\delta(x, z) \subseteq \delta(x, y) \cup \delta(y, z)$.

These properties resemble properties of semi-distance, and therefore the function δ may be regarded as *qualitative semi-metric* and $\delta(x, y)$ – *qualitative semi-distance*. Thus the discernibility matrix can be seen as a *semi-distance* (*qualitative*) matrix.

Let us also note that for every $x, y, z \in U$ we have

iv) $|\delta(x, x)| = 0$,

v) $|\delta(x, y)| = |\delta(y, x)|$,

vi) $|\delta(x, z)| \leq |\delta(x, y)| + |\delta(y, z)|$.

In order to compute D-reducts of condition attributes C, we will need slightly modified discernibility matrix, called the (C, D)-matrix, which is given next

$$\delta(x, y) = \{a \in C: a(x) \neq a(y) \text{ and } w(x, y)\},$$

where $w(x, y) \equiv x \in POS_C(D)$ and $y \notin POS_C(D)$ or

$x \notin POS_C(D)$ and $y \in POS_C(D)$ or

$x, y \in POS_C(D)$ and $(x, y) \notin I_D$

for $x, y \in U$.

If the partition defined by D is definable by C then the condition $w(x, y)$ in the above definition can be reduced to $(x, y) \notin I_D$.

Thus entry $\delta(x, y)$ is the set of all attributes which discern objects x and y that do not belong to the same equivalence class of the relation I_D.

The (C, D)-matrix for Table 1 with condition attributes C, S, P and decision attribute *cracks* is given in Table 5.

	1	2	3	4	5	6
1						
2	C					
3	–	$\varnothing$				
4	C,S	C,S	C,S			
5	C,S,P	–	S,P			
6	–	C,S,P	–	C,P	C	

Table 5

Set $C' \subseteq C$ is the D-reduct of C, if C' is the minimal (with respect to inclusion) subset of C such that

$$C' \cap c \neq \varnothing \text{ for any nonempty entry } c(c \neq \varnothing) \text{ of the } (C,D) \text{ - discernibility matrix.}$$

Thus D-reduct is the minimal subset of attributes that discerns all equivalence classes of the relation I_D discernible by the whole set of attributes. Every (C, D)-discernibility matrix defines uniquely a *discernibility* (*Boolean*) *function* $f_D(C)$ defined as follows.

Let us assign to each attribute a a binary Boolean variable, $\bar{a}$ and let $\Sigma\, \delta(x, y)$ denotes Boolean sum of all Boolean variables assigned to the set of attributes $\delta(x, y)$. Then the discernibility function can be defined by the formula

$$f_D(C) = \prod_{(x,y)\in U^2} \{\Sigma\delta(x,y):(x,y) \in U^2 \text{ and } \delta(x,y) \neq \varnothing\}.$$

The following property establishes the relationship between disjunctive normal form of the function $f_D(C)$ and the set of all D-reducts of C.

All constituents in the minimal disjunctive normal form of the function $f_D(C)$ *are all D-reducts of C.*

In other words, a reduct is a minimal subset of attributes that discerns all objects discernible by the whole set of attributes.

The (C, P)-discernible function for the (C, D)-discernible matrix shown in Table 5 is the following

$$f_D(C) = C \cdot (C + S) \cdot (C + S + P) \cdot (C + S) \cdot (C + S + P) \cdot (C + S) \cdot (S + P) \cdot C = C \cdot (S + P),$$

where „+” and „·”denote Boolean addition and multiplication, respectively. Because the disjunctive normal from of the function is

$$f_D(C) = C \cdot S + C \cdot P$$

hence we have two D-reducts $\{C, \mathrm{S}\}$ and $\{C, P\}$ of the set of condition attributes $\{C, S, P\}$.
Let us also observe that the D-core is the set of all single element entries of the discernibility matrix, i.e.,

$$CORE_D(C) = \{a \in C: \delta(x,y) = \{a\}, \text{ for some } x, y\}.$$

For very large data tables the proposed method of computing reducts is not efficient enough and more sophisticated approaches are used. For details the reader is advised to consult the references.

SIGNIFICANCE OF ATTRIBUTES

As it follows from considerations concerning reduction of attributes, they can be not equally important, and some of them can be eliminated from an information table without loosing information contained in the table. The idea of attribute reduction can be generalized by introduction a concept of *significance of attributes*, which enables us to evaluate of attributes not only by two-valued scale, *dispensable – indispensable*, but by assigning to an attribute a real number from the closed interval [0,1], expressing how important is an attribute in an information table.

Significance of an attribute can be evaluated by measuring effect of removing the attribute from an information table on classification defined by the table. Let C and D be sets of condition and decision attributes respectively and let a be a condition attribute. As shown previously the number $\gamma(C, D)$ expresses the degree of dependency between attributes C and D. We can ask how the coefficient $\gamma(C, D)$ changes when removing an attribute a, i.e., what is the difference between $\gamma(C, D)$ and $\gamma((C - \{a\}, D)$. We can normalize the difference and define the significance of an attribute a as

$$\sigma_{CD}(a) = \frac{(\gamma(C,D) - \gamma(C - \{a\}, D))}{\gamma(C,D)} = 1 - \frac{\gamma(C - \{a\}, D)}{\gamma(C,D)},$$

and denoted simple by $\sigma(a)$, if C and D are understood.

Obviously $0 \le \sigma(a) \le 1$. The more important is the attribute a the greater is the number $\sigma(a)$.

For example for condition attributes in Table 1 we have the following results:

$\sigma(C) = 0{,}75$,
$\sigma(S) = 0{,}00$,
$\sigma(P) = 0{,}00$.

Because the significance of the attributes S and P are zero, removing either of the attribute from condition attributes does not effect the dependency, whatsoever. Hence the attribute C is the most significant one in the table. That means that by removing the attribute C, 75% (three out of four) of pipes cannot be properly classified.

However for reduced date table, e.g., Table 2 we get

$\sigma(C) = 1{,}00$,
$\sigma(S) = 0{,}25$.

In this case, removing the attribute S from the reduct, i.e., using only the attribute C, 25% (one out of four) objects can be properly classified, while dropping the attribute C, i.e., using only the attribute S, 100 % objects (all) cannot be classified.

That means that in this case making decisions is impossible at all, whereas by employing only the attribute C some decision can be made.

Thus the coefficient $\sigma(a)$ can be understood as an error which occurs when attribute a is dropped. The significance coefficient can be extended to set of attributes as follows:

$$\sigma_{CD}(B) = \frac{(\gamma(C,D) - \gamma(C - B, D))}{\gamma(C,D)} = 1 - \frac{\gamma(C - B, D)}{\gamma(C,D)}$$

denoted by $\sigma(B)$, if C and D are understood, where B is a subset of C.

If B is a reduct of C, then $\sigma(B) = 1$, i.e., removing any reduct from a set of decision rules unables to make decisions with certainty, whatsoever.

Any subset B of C will be called an *approximate reduct* of C, and the number

$$\varepsilon_{CD}(B) = \frac{(\gamma(C,D) - \gamma(C-B,D))}{\gamma(C,D)} = 1 - \frac{\gamma(C-B,D)}{\gamma(C,D)}$$

denoted simply as $\varepsilon(B)$, will be called an *error of reduct approximation*. It expresses how exactly the set of attributes B approximates the set of condition attributes C. Obviously $\varepsilon(B) = 1 - \sigma(B)$ and $\varepsilon(B) = 1 - \varepsilon(C - B)$.

For any subset B of C we have $\varepsilon(B) \leq \varepsilon(C)$. If B is a reduct of C, then $\varepsilon(B) = 0$.

For example, either of attributes S and C can be considered as approximate reducts of $\{C, S\}$, and

$\varepsilon(C) = 1{,}00$.

But for the whole set of condition attributes $\{C, S, P\}$ we have also the following approximate reduct

$\varepsilon(S, P) = 0.75$.

The concept of an approximate reduct is a generalization of the concept of a reduct considered previously. A minimal subset B of condition attributes C, such that $\gamma(C,D) = \gamma(B,D)$, or $\varepsilon_{(C,D)}(B) = 0$ is a reduct in the previous sense.

The idea of an approximate reduct can be useful in cases when a smaller number of condition attributes is preferred over accuracy of classification.

DECISION RULES AND DEPENDENCY

With every dependency $C \Rightarrow_k D$ we can associate a set of decision rules, specifying decisions that should be taken when certain condition are satisfied. In other words every decision table determines a set of formulas of the form:

„*if* ... *then*".

For example, Table 1 determines the following set of decision rules:

(1) *if* (*C*, *high*) *and* (*S*, *high*) *and* (*P*, *low*) *then* (*Cracks*, *yes*),
(2) *if* (*C*, *avg*.) *and* (*S*, *high*) *and* (*P*, *low*) *then* (*Cracks*, *no*),
(3) *if* (*C*, *avg*.) *and* (*S*, *high*) *and* (*P*, *low*) *then* (*Cracks*, *yes*),
(4) *if* (*C*, *low*) *and* (*S*, *low*) *and* (*P*, *low*) *then* (*Cracks*, *no*),
(5) *if* (*C*, *avg*.) *and* (*S*, *low*) *and* (*P*, *high*) *then* (*Cracks*, *no*),
(6) *if* (*C*, *high*) *and* (*S*, *low*) *and* (*P*, *high*) *then* (*Cracks*, *yes*).

From logical point of view decision rules are implications built up form elementary formulas of the form (attribute name, attribute value) and combined together by means of proportional connectives „and", „or" and „implication" in a usual way.

Predecessor of the implication specifies conditions, which should be fulfilled in order to perform decisions determined by the successor of the implication.

Decision rule is *deterministic* (*certain*, *sure*) if its conditions uniquely determine decisions, otherwise the decision rule is *nondeterministic* (*uncertain*, *possible*).
In the example above rules (1) , (4), (5) and (6) are deterministic, whereas rule (2) and (3) are nondeterministic.

Obviously only deterministic rules determine unambiguous decisions.

Because decision rules are logical formulas they can be simplified using standard logical methods not presented here. Besides, they can be also simplified employing rough set approach. For example, the concept of the reduct leads to elimination of superfluous condition attributes. Hence, instead of Table 1 we can use Table 2 or Table 3 to obtain simplified decision rules.

For Table 2 we obtain

(1) *if* (*C*, *high*) *and* (*S*, *high*) *then* (*Cracks*, *yes*),
(2) *if* (*C*, *avg.*) *and* (*S*, *high*) *then* (*Cracks*, *no*),
(3) *if* (*C*, *avg.*) *and* (*S*, *high*) *then* (*Cracks*, *yes*),
(4) *if* (*C*, *low*) *and* (*S*, *low*) *then* (*Cracks*, *no*),
(5) *if* (*C*, *avg.*) *and* (*S*, *low*) *then* (*Cracks*, *no*),
(6) *if* (*C*, *high*) *and* (*S*, *low*) *then* (*Cracks*, *yes*),

and for Table 3 we have

(1'') *if* (*C*, *high*) *and* (*P*, *low*) *then* (*Cracks*, *yes*),
(2'') *if* (*C*, *avg.*) *and* (*P*, *low*) *then* (*Cracks*, *no*),
(3'') *if* (*C*, *avg.*) *and* (*P*, *low*) *then* (*Cracks*, *yes*),
(4'') *if* (*C*, *low*) *and* (*P*, *low*) *then* (*Cracks*, *no*),
(5'') *if* (*C*, *avg.*) *and* (*P*, *high*) *then* (*Cracks*, *no*),
(6'') *if* (*C*, *high*) *and* (*P*, *high*) *then* (*Cracks*, *yes*).

Using the rough set techniques, not presented in this paper, the decision rules can be simplified further, e.g., for Table 2 we get

(1''') *if* (*C*, *high*) *then* (*Cracks*, *yes*),
(2''') *if* (*C*, *avg.*) *and* (*S*, *high*) *then* (*Cracks*, *no*),
(3''') *if* (*C*, *avg.*) *and* (*S*, *high*) *then* (*Cracks*, *yes*),
(4''') *if* (*C*, *low*) *then* (*Cracks*, *no*),
(5''') *if* (*C*, *avg.*) *and* (*S*, *low*) *then* (*Cracks*, *no*),
(6''') *if* (*C*, *high*) *and* (*S*, *low*) *then* (*Cracks*, *yes*).

In order to express certainty of decision specified by a decision rule we would need numerical characterization of the rule, showing to what extend the decision can be trusted. To this end we define a *certainty factor* of the rule.

Let Φ and Ψ be logical formulas representing conditions and decisions, respectively and let $\Phi \to \Psi$ be a decision rule, where. Φ_S denote the meaning of Φ in the system *S*, i.e., the set of all objects satisfying Φ in *S*, defined in a usual way.

With every decision rule $\Phi \to \Psi$ we associate a number, called the *certainty factor* of the rule, and defined as

$$\mu(\Phi, \Psi) = \frac{|\Phi_S \cap \Psi_S|}{|\Phi_S|}.$$

Of course $0 \le \mu(\Phi, \psi) \le 1$; get the rule $\Phi \to \Psi$ is deterministic then $\mu(\Phi, \psi) = 1$, and for nondeterministic rules $\mu(\Phi, \psi) < 1$.

For example, the certainty factor for decision rules consider previously are as follows:

$$\mu(\Phi_1, \Psi_1) = 1, \quad \mu(\Phi_4, \Psi_4) = 1,$$
$$\mu(\Phi_2, \Psi_2) = 1/2, \quad \mu(\Phi_5, \Psi_5) = 1,$$
$$\mu(\Phi_3, \Psi_3) = 1/2, \quad \mu(\Phi_6, \Psi_6) = 1,$$

where Φ_i, Ψ_i denote conditions and decisions of the rule *i*.

Let us notice that the certainty factor can be viewed as a generalization of the rough membership function and can be also interpreted as a conditional probability that an object *x* satisfies decision provided it satisfies condition of the rule.

In fact decision rules obtained from a data table are logical inference rules which allow to discover patterns hidden in data, and can be used to reason about reality, or – as a starting point for implementation of computer support decision systems, expert systems, control algorithms etc.

Hence with every decision rule $\Phi \rightarrow \Psi$ one can associate a rough deduction rule called *rough modus ponens*

$$\frac{\pi(\Phi);\ \mu(\Phi,\Psi)}{\pi(\Psi)},$$

where $\pi(\Phi) = \frac{|\Phi_S|}{|U|}$ and $\pi(\Psi) = \pi(\sim\Phi \wedge \Psi) + \pi(\Phi) \cdot \mu(\Phi,\Psi)$.

This rule allows to compute the probability of decisions in terms of probability of conditions and conditional probability of the decision rule.

For example, it is easy to compute that if Φ = (*C*, *avg*.), (*S*, *high*) (*P*, *low*) and Ψ = (*Cracks*, *yes*) then we have $\pi(\Phi)$ = 1/3, $\mu(\Phi, \Psi)$ = 1/2 and $\pi(\Psi)$ = 1/3; for Φ = (*C*, *high*) and Ψ = (*Cracks*, *yes*) we get $\pi(\Phi)$ = 1/3, $\mu(\Phi, \Psi)$ = 1 and $\pi(\Psi)$ = 1/3, whereas for Φ = (*S*, *low*) and Ψ = (*Cracks*, *yes*) we obtain $\pi(\Phi)$ = 1/2, $\mu(\Phi, \Psi)$ = 1/6 and $\pi(\Psi)$ = 1/4.

The problem considered above is a part of a wider question pursued for many years in AI and is related to common-sense reasoning methods. In classical logic basic rule of inference is grounded on the assumption that if a premise Φ and the implication $\Phi \rightarrow \Psi$ are true then the conclusion Ψ must be also true. This deduction rule is known as *modus ponens*. However in the common-sense reasoning methods we must admit that a premise and a inference rule are often not known with certainty, but with some probability and therefore the conclusion must be also equipped with a proper probability measure. Classical logic does not offer methods to solve this dilemma and paradigm of classical logic is no more valid in this case. Consequently *modus ponens* cannot be postulated as a fundamental reasoning rule for common-sense reasoning.

Very many mathematical models have been proposed to solve this problem and master uncertainty in reasoning. The presented rough set approach seems to be a very natural answer to this problem and it has a very inherent interpretation in data sets

APPLICATIONS METHODOLOGY

Rough set theory is mainly meant as a new mathematical approach to discover patterns in data, i.e., we are given a data set as a result of observations of some real-life phenomena and our task is to find out hidden patterns in the data. The patterns are usually represented in a form of a set of decision rules. The rules can be used to explain the data, i.e., to explain the phenomena or processes underlying the data and are usually interpreted as a description of some „cause-effect" relations. The example of the endurance test of cast iron pipes, given in the introduction, is a very good illustrative example of this kind of applications.

However there are many other possibilities of application of rough set theory. Beside analysis of data specially interesting seems the rough set approach to processes specification, which will be discussed briefly next, by means of a simple example of a distributed traffic signal control.

Let us consider a very simple intersection (T-intersection) shown in Fig 1.

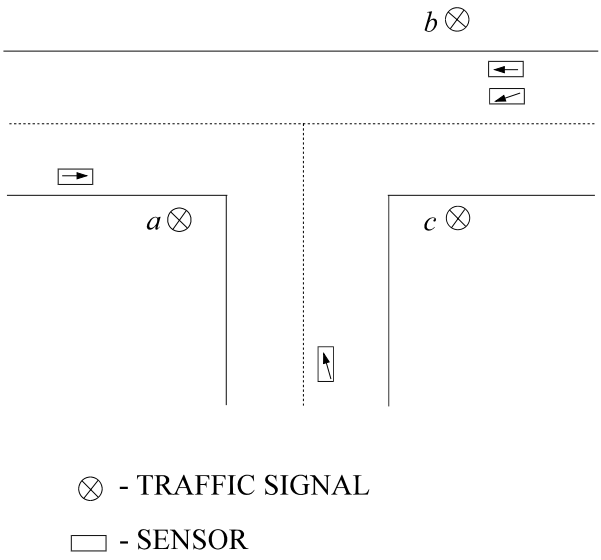

Fig. 1

We would like to design a distributed control algorithm which will supervise the traffic on the basis of local conditions. We assume that the conditions are determined by sensors placed in lanes and indicating the desired turn of a car approaching the intersection. For the sake of simplicity we omit many important factors, needed in real-life control and assume that the specification of the control admissible is situations is given in Table 6 below

	a	*b*	*c*
1	1	1	0
2	0	2	0
3	0	0	2

Table 6

where

0 – red
1 – green
2 – green arrow (left turn)

Let us observe that Table 6 is not a result of observations but, a specification of requirements for a special discrete systems behavior (e.g., controller, program etc.).

Using methods shown previously we get from the table the following control rules

if (*b*, 0) *or* (*b*, 2) *then* (*a*, 0)
if (*b*, 1) *then* (*a*, 1)
if (*c*, 2) *then* (*b*, 0}
if (*a*, 1) *then* (*b*, 1}
if (*a*, 0) *and* (*c*, 0) *then* (*b*, 2)
if (*b*, 1) *or* (*b*, 2) *then* (*c*, 0)
if (*b*, 0) *then* (*c*, 2)

The corresponding switching circuits for the set of control rules is shown in Fig. 2

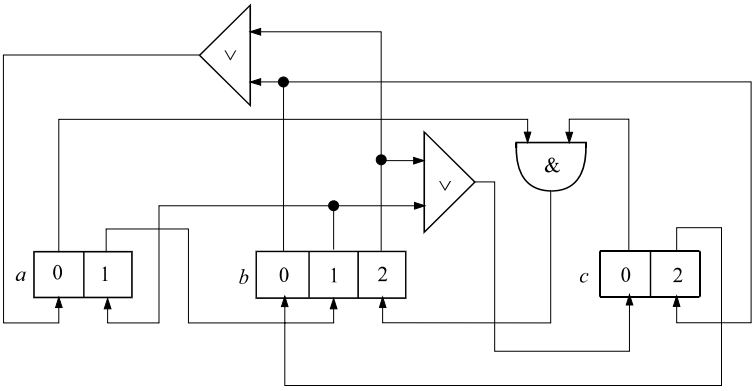

Fig. 2

and the corresponding controller is depicted in Fig. 3

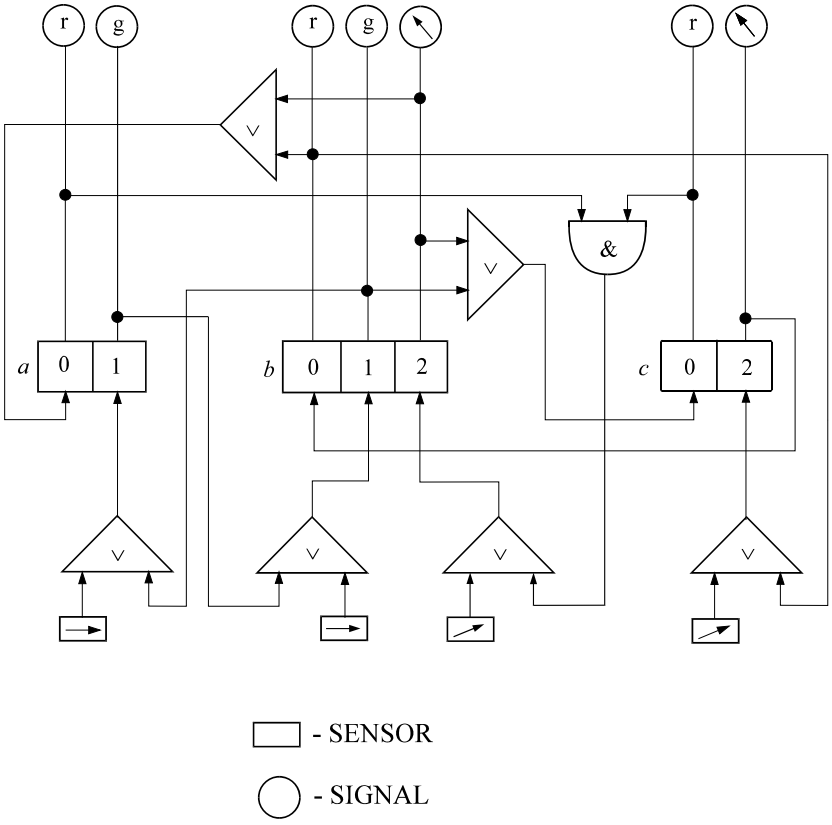

Fig. 3

In the example above in fact we specify by means of rough set techniques some kind of concurrent process. This class of processes play a very important role in many domains, in particular in industrial environment, and no doubt, the most successful formal model of concurrent processes are Petri Nets. Rough set methodology can be also useful to specification of concurrent processes and synthesis of complex control systems. But this area of applications need more research. Some results in this domain can be found in many research papers.

CONCLUSION

Rough set approach to data analysis has many important advantages. Some of them are listed below.

- Provides efficient algorithms for finding hidden patterns in data.
- Identifies relationships that would not be found using statistical methods.
- Allows both qualitative and quantitative data.
- Finds minimal sets of data (data reduction).
- Evaluates significance of data.
- Generates 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 computer organization based on rough set theory is necessary.

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

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

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 new discretization methods for quantitative attribute values are badly needed. Also an extensive study of a new approach to missing data is very important. 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 advanced applications. Some research in this area is already in progress.

For basic ideas of rough set theory the reader is referred to [8,9,15,18].

ACKNOWLEDGMENTS

The author gratefully acknowledge the support of the Air Force Contract F61708-97-WO196.

REFERENCES

1. J.W. Grzymała-Busse, in Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory, (R. Słowiński ed.), Kluwer Academic Publishers, Boston, London, Dordrecht, 1992, P.471.
2. A.G. Jackson, M. Ohmer and H. Al-Kamhawi, in The Third International Workshop on Rough Sets and Soft Computing Proceedings (RSSC'94), (T.Y. Lin ed.), San Jose State University, San Jose, California, USA, 1994.
3. A.G. Jackson, S.R. LeClair, M.C. Ohmer, W. Ziarko and H. Al-Kamhwi, Acta Metallurgica et Materialia, 1996, p.4475.

4. E. Krusińska, R. Słowiński and J. Stefanowski, Applied Stochastic Models and Data Analysis, 8, 1992, p.43.
5. T.Y. Lin and N. Cercone, Rough Sets and Data Mining - Analysis of Imperfect Data, Kluwer Academic Publishers, Boston, London, Dordrecht, 1997, P.430.
6. T.Y. Lin and A.M.Wildberger, The Third International Workshop on Rough Sets and Soft Computing Proceedings RSSC'94), San Jose State University, San Jose, California, USA, 1995.
7. Z. Pawlak, International Journal of Computer and Information Sciences, 11, 1982, p.341.
8. Z. Pawlak, Rough Sets - Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Boston, London, Dordrecht, 1991, P.229.
9. Z. Pawlak, J.W. Grzymała-Busse, R. Słowiński and W. Ziarko, Communication of the ACM, 38, 1995, p.88.
10. Z. Pawlak and A. Skowron, in Advances in the Dempster Shafer Theory of Evidence, (R.R Yaeger, M. Fedrizzi and J. Kacprzyk eds.), John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore, 1994.
11. Z. Pawlak and R. Słowiński, European Journal of Operational Research, 72, 1994, p.443.
12. A. Skowron and J.W. Grzymała-Busse, in Advances in the Dempster-Shafer Theory of Evidence, (R.R, Yaeger, M. Fedrizzi and J. Kacprzyk eds.), John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore, 1994.
13. A. Skowron and C. Rauszer, in Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory, (R. Słowiński ed.), Kluwer Academic Publishers, Boston, London, Dordrecht, 1992, P.471.
14. R. Słowiński, Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory, Kluwer Academic Publishers, Boston, London, Dordrecht, 1992, P.471.
15. R. Słowiński, AI Expert, 10, 1995, p.18.
16. R. Słowiński and J. Stefanowski, J., (1992), in Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory, (R. Słowiński ed.), Kluwer Academic Publishers, Boston, London, Dordrecht, 1992, P.471.
17. A. Szladow, PC AI , 7/1, 1993, p.40.
18. A. Szladow and W. Ziarko, AI Expert, 7,1993, p.36.
19. S. Tsumoto, S. Kobayashi, T. Yokomori, H. Tanaka and A. Nakamura, The Fourth Internal Workshop on Rough Sets, Fuzzy Sets and Machine Discovery, PROCEEDINGS, The University of Tokyo, 1996, P.465.
20. P.P. Wang, Second Annual Joint Conference on Information Sciences, PROCEEDINGS, Wrightsville Beach, North Carolina, USA, 1995.
21. P. Wang, Joint Conference of Information Sciences, Vol. 3. Rough Sets and Computer Sciences, Duke University, 1997, P.449.
22. W. Ziarko, Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery (RSKD'93), Banff, Alberta, Canada, October 12--15, Springer-Verlag, Berlin, 1993, P.476.