Rough Sets: An Approach to Vagueness

Zdzisław Pawlak
Institute of Theoretical and Applied Informatics, Polish Academy of Sciences
Bałtycka 5, 44-100 Gliwice, Poland
and
Warsaw School of Information Technology
Newelska 6, 01-447 Warsaw, Poland

Lech Polkowski
Polish--Japanese Institute of Information Technology
Koszykowa 86, 02-008 Warsaw, Poland
and
Department of Mathematics and Computer Science
University of Warmia and Mazury
Żołnierska 14a, 10-561 Olsztyn, Poland

Andrzej Skowron
Institute of Mathematics, Warsaw University
Banacha 2, 02-097 Warsaw, Poland

ROUGH SETS AS A TOOL FOR REASONING ABOUT VAGUE CONCEPTS

Rough set theory is a new mathematical approach to imperfect knowledge. The problem of imperfect knowledge, tackled for a long time by philosophers, logicians and mathematicians, has become 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 imperfect knowledge. The most successful one is, no doubt, fuzzy set theory proposed by Zadeh (Zadeh, 1965). Rough set theory (Pawlak, 1982) presents still another attempt at this problem. This theory has attracted attention of many researchers and practitioners all over the world, who contributed essentially to its development and 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.

The rough set approach provides efficient algorithms for finding hidden patterns in data, minimal sets of data (data reduction), evaluating significance of data, generating sets of decision rules from data. This approach is easy to understand, offers straightforward interpretation of obtained results, most of its algorithms are particularly suited for parallel processing.

The rough set philosophy is founded on the assumption that with every object of the universe of discourse we associate some information. For example, 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 rough set theory.

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

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

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 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 possibly belong to the concept. The difference between the upper and the lower approximation constitutes *the boundary region* of the vague concept. Approximations are two basic operations in rough set theory. The observation that vague concepts should have non-empty boundary was made by Gottlob Frege (Frege, 1905).

Vague complex concept approximation and reasoning about vague concepts by means of such approximations become critical for numerous applications related to multiagent systems (such as web mining, e-commerce, monitoring, security and rescue tasks in multiagent systems, cooperative problem solving, intelligent smart sensor fusion, human-computer interfaces, telemedicine, soft views of databases for specific customers). Further development of such methods for approximate reasoning by intelligent agents (Russell & Norvig, 2003) based on databases and knowledge bases is needed.
Let us consider essential issues in rough sets in more detail.

Table 1: A summary of basic research ideas of rough set theory

Data tables. Data for rough set based analysis are usually formatted into a data table (an *information system*) **A**= (U,A), where the set U consists of *objects* (e.g., records, signals, processes, patients) and the set A consists of *attributes* (e.g., physical parameters, features expressed in symbolic or numerical form, results of medical tests); any attribute $a\in A$ is a mapping from U into a value set V_a. Subsets of U are *concepts*.

Indiscernibility relation. The A-*indiscernibility relation*, IND(A) is defined as follows, x IND(A) y iff $a(x)=a(y)$ for $a\in A$. For $B \subseteq A$, one may consider a restricted information system A(B) = (U,B) and define the B-indiscernibility IND(B). For a set B of attributes, we denote by $[x]_B$ the equivalence class of $x\in U$ with respect to IND(B). In terms of indiscernibility, important notions related to knowledge reduction and attribute dependence are expressed.

Reducts. For a set B of attributes, one can look after an inclusion-minimal set $C \subseteq B$ with the property that IND(C)=IND(B), i.e., C is a minimal subset of attributes in B that provides the same classification of concepts as B. Such C is said to be a B-*reduct*. The problem of finding a minimum-length reduct is NP-hard (Skowron & Rauszer, 1992), so heuristics are used in searching for short (or relevant with respect to a give criterion) reducts. We mention heuristics based on the Johnson algorithm or genetic algorithms (Bazan et al, 1998). Given a B-reduct C, one can reduce the information system A(B) to the system A(C) without any loss of classification ability. Many other kinds of reducts and their approximations are discussed in literature. They are used in searching for relevant patterns in data (Polkowski & Skowron, 1998; Polkowski, Tsumoto & Lin, 2002).

Functional dependence. For given A= (U,A), $C,D \subseteq A$, by $C\rightarrow D$ is denoted the *functional dependence* of D on C in A that holds iff $IND(C) \subseteq IND(D)$. In particular, any B-reduct C determines functionally D. Also dependencies to a degree are considered (Pawlak, 1991).

Definable and rough concepts (sets). Classes of the form $[x]_B$ can be regarded as the primitive B-*definable concepts w*hose elements are classified with certainty by means of attributes in B. This property extends to more general concepts, i.e., a concept $X \subseteq U$, is B-*definable* iff for each y in U, either $[y]_B \subseteq U$ or $[y]_B \cap X = \emptyset$. This implies that X has to be the union of a collection of B-indiscernibility classes, i.e., $X = \cup\{[x]_B: x\in X\}$. Then we call X a B-*exact* (*crisp*, *precise*) concept. One observes that unions, intersections and complements in U to B-exact concepts are B-exact as well, i.e., B-exact concepts form a Boolean algebra for each $B\subseteq A$. In case when a concept X is not B-exact, it is called B-*rough*, and then X is described by *approximations* of X that are exact concepts (Pawlak, 1991), i.e., one defines the B-*lower approximation* of X, and the B-*upper approximation* of X by $B_*(Y) = \{x\in X: [x]_B \subseteq X\}$ and $B^*(Y) =\{x\in X: [x]_B \cap X\neq\emptyset\}$, respectively. The set $B^*(Y) - B_*(Y)$ is called the B-*boundary region* of X.

Rough membership functions. Roughness of Y with respect to a set B can be measured by some coefficients (Pawlak, 1991).
A precise local characteristics of a concept X with respect to a classification IND(B) in an information system (U,A) can be given

by a *rough membership function* (Pawlak & Skowron, 1994), defined as the fraction of the class $[x]_B$ contained in X. This definition is relative to a given source of information what makes the rough membership function different from fuzzy membership function. Notice that such a coefficient has been considered by Łukasiewicz (Łukasiewicz, 1913) long time ago in studies on assigning fractional truth values to logical formulas.

Decision systems and rules. Matching classification of objects by an expert with a classification in terms of accessible features, can be done with decision systems. A *decision system* is a tuple $A^d=(U,A,d)$, where (U,A) is an information system with the set A of condition attributes, and the decision (attribute) $d: U \rightarrow V_d$, where $d \notin A$. In case $A \rightarrow d$ holds in A^d, we say that the decision system A^d is deterministic and the dependency $A \rightarrow d$ is A^d-exact. Then, for each class $[x]_A$ there exists a unique decision d(x) throughout the class. Otherwise, the dependency $A \rightarrow d$ in A^d holds to a degree. A *decision rule* in A^d is any expression $\wedge\{a=v_a : a \in A \text{ and } v_a \in V_a\} \rightarrow d=v$ where d is the decision attribute and $v \in V_d$. This decision rule is true in (U,A,d) if for any object satisfying its left hand side it also satisfies the right hand side, otherwise the decision rule is true to a degree measured by some coefficients (Pawlak, 1991). Strategies for inducing decision rules can be found in (Polkowski & Skowron, 1998; Polkowski, Tsumoto & Lin, 2000).

Approximation spaces. Several generalizations of the classical rough set approach based on approximation spaces have been reported in the literature. Let us consider some examples. A *generalized approximation space* is defined by a tuple $AS=(U,I,\nu)$ where I is the *uncertainty function* defined on U with values in the powerset P(U) of U (I(x) is the neighboorhood of x) and ν is the *inclusion function* (called also *rough inclusion*) defined on the Cartesian product $P(U) \times P(U)$ with values in the interval [0,1] measuring the degree of inclusion of sets. The standard rough inclusion is defined by $\nu(X,Y) = |X \cap Y|/|X|$ if X is nonempty, and 1 otherwise, for $X,Y \subseteq U$. The lower AS_* and upper AS^* approximation operations can be defined in AS by $AS_*(X)=\{x \in U: \nu(I(x),X)=1\}$ and $AS^*(X)=\{x \in U: \nu(I(x),X) > 0\}$, respectively. In the standard case I(x) is equal to the equivalence class $[x]_B$ of the indiscernibility relation IND(B); in case of tolerance (similarity) relation $\tau \subseteq U \times U$ we take $I(x)=\{y \in U: x\tau y\}$, i.e., I(x) is equal to the tolerance class of τ defined by x. Usually, there are considered families of approximation spaces with approximation spaces labeled by some parameters. By tuning such parameters according to chosen criteria (e.g., the minimal description length principle) one can search for the optimal approximation space for concept description.

Rough mereology. The approach based on inclusion functions was generalized to the *rough mereological approach* (Polkowski & Skowron 1996; Polkowski, Tsumoto & Lin, 2000; Pal, Polkowski & Skowron, 2004). The inclusion relation $x\mu_r y$ with the intended meaning "x *is a part of y to a degree at least* r" has been taken as the basic notion of the *rough mereology* that is a generalization of the Leśniewski mereology. Rough mereology offers a methodology for synthesis and analysis of complex objects in distributed environment of intelligent agents, in particular, for synthesis of objects satisfying a given specification to a satisfactory degree or for control in such complex environment. Moreover, rough mereology has been recently used for developing foundations of the *information granule calculi* (Pal, Polkowski & Skowron, 2004), aiming at formalization of the Computing with Words and Perceptions paradigm, recently formulated in (Zadeh, 2001). More complex information granules are defined recursively using already defined information granules and their measures of *inclusion* and *closeness*. Information granules such as classifiers (Duda, Hart & Stork, 2002) or approximation spaces can have complex structures. Computations on information granules are performed to discover relevant information granules, e.g., patterns or approximation spaces for complex concept approximations.

There are also developed extensions of rough sets to deal with concept approximations using inductive reasoning.

Tasks collected under labels of data mining, knowledge discovery, decision support, pattern classification, approximate reasoning require tools aimed at discovering in data of templates (patterns) and classifying them into certain decision classes. Templates are in many cases most frequent sequences of events, most probable events, regular configurations of objects, the decision rules of high quality, approximate reasoning schemes. Tools for discovering and classifying of templates are based on reasoning schemes rooted in various paradigms (Duda, Hart & Stork, 2002; Kloesgen & Żytkow, 2002). Such patterns can be extracted from data by means of methods based on Boolean reasoning and discernibility. The *discernibility relation* (in the simplest case defined as the complement of the indicernibility relation) is one of the most important relation considered in rough set theory. The ability to discern between perceived objects is important for constructing many entities like reducts, decision rules or decision algorithms. The idea of *Boolean reasoning* is based on construction for a given problem P of a corresponding Boolean function f_P with the following property: the solutions for the problem P can be decoded from prime implicants of the Boolean function f_P. Let us mention that to solve real-life problems it is necessary to deal with Boolean functions having large number of variables.

A list of some current research directions on the rough set foundations and the rough set based methods is presented in Table 2. For more details the reader is referred to the enclosed bibliography on rough sets (Pawlak, 1991; Słowiński 1992; Orłowska 1997; Lin & Cercone, 1997; Polkowski & Skowron, 1998; Pal & Skowron, 1999; Polkowski, Tsumoto & Lin, 2000; Polkowski, 2002; Demri & Orłowska, 2002; Pal, Polkowski & Skowron, 2004).

Table 2: A List of Research Directions on Rough Sets

- Boolean reasoning and approximate Boolean reasoning strategies as the basis for efficient heuristics for rough set methods.
- Tolerance (similarity) based rough set approach.
- Rough set based approach based on neighborhood (uncertainty) functions and inclusion relation. In particular, variable precision rough set model.
- Rough sets in multi-criteria decision analysis and preference modeling.
- Rough sets and non-deterministic information systems.
- Rough set based clustering.
- Rough sets and incomplete information systems. In particular, missing value problems.
- Rough sets and noisy data.
- Rough sets and relational databases.
- Rough sets and inductive reasoning.
- Rough sets in modeling of decision systems and analysis of complex systems, in particular, rough sets and layered (hierarchical) learning.
- Rough sets as a tool for approximate reasoning in distributed systems, by autonomous agents, and in multi—agent systems.
- Calculi of information granules.

Table 3: An assessment of results and challenges

- A successful methodology based on the discernibility of objects and Boolean reasoning was developed for computing of many different kinds of reducts and their approximations that are used (i) for inducing, e.g., decision rules, association rules, discretization of real value attributes, symbolic value grouping, (ii) in searching for new features defined by oblique hyperplanes, higher order surfaces, and relevant patterns from data, (iii) in conflict resolution or negotiation.
- Most of the problems related to generation of the above reducts are NP-complete or NP-hard. However, it was possible to develop efficient heuristics returning suboptimal solutions of the problems. The results of experiments on many data sets are very promising. They show very good quality of solutions generated by the heuristics in comparison with other methods reported in literature (e.g., with respect to the classification quality of unseen objects). Moreover, they are very efficient from the point of view of time necessary for computing of the solution. It is important to note that the methodology makes it possible to construct heuristics having a very important approximation property: expressions generated by heuristics (i.e., implicants) close to prime implicants define approximate solutions for the problem (Polkowski & Skowron, 1998).
- A wide range of applications of methods based on rough set theory alone or in combination with other approaches have been discovered in the following areas (Polkowski & Skowron, 1998; Pal, Polkowski & Skowron, 2004): acoustics, biology, business and finance, chemistry, computer engineering (e.g., data compression, digital image processing, digital signal processing, parallel and distributed computer systems, sensor fusion, fractal engineering), decision analysis and systems, economics, electrical engineering (e.g., control, signal analysis, power systems), environmental studies, digital image processing, informatics, medicine, molecular biology, musicology, neurology, robotics, social science, software engineering, spatial visualization, Web engineering, and Web mining.
- Several software systems based on rough sets have been developed such as ROSETTA or RSES (see references).
- Rough sets in combination with other soft and computing technologies such as fuzzy sets, evolutionary programming, neural networks or crisp technologies offered, e.g., by statistical or analytical techniques are promising in solving some tasks however they should be further developed to deal with hard real-life problems.

F U T U R E P L A N S

To enhance the mentioned above methods (see Table 3) further development of methods based on rough sets in combination with other soft computing techniques is needed. In particular, methods based on rough mereological approach (Polkowski & Skowron, 1996) combined with information granulation (Pal, Polkowski & Skowron, 2004) seems to be very promising. Among issues to be investigated are scalability problems, methods for designing adaptive systems, methods for intelligent system constructing that perform computations with words and perceptions (Zadeh, 2001).

C O N C L U S I O N

Rough set theory supplies essential tools for knowledge analysis. It allows for creating algorithms for knowledge reduction, concept approximation, decision rule induction, and object classification. The methods of rough set theory rest on indiscernibility and related notions, in particular on notions related to rough inclusions. All constructs needed in implementing rough set based algorithms can be derived from data tables with need for neither a priori estimates nor preliminary assumptions. Currently, research in rough set theory is directed, among others, at problems of knowledge granulation, computing with words and perceptions techniques and rough–neural computing (Pal, Polkowski & Skowron, 2004).

R E F E R E N C E S

Bazan, J., Nguyen, H.S., Nguyen, S.H., Synak, P. and Wróblewski, J. (2000). Rough set algorithms in classification problems, *Rough Set Methods and Applications. New Developments in Knowledge Discovery in Information Systems*, Polkowski, L., Tsumoto, S. and Lin, T.Y. (eds.), Heidelberg: Physica-Verlag, 49-88.

Demri, S. and Orłowska, E. (2002). *Incomplete Information: Structure, Inference, Complexity*, Springer-Verlag, Heidelberg.

Duda, R., Hart, P. and Strok, R. (2002). *Pattern Classification,* New York: Wiley.

Frege, G. (1903). *Grundgesetzen der Arithmetik*, 2. Jena: Verlag von Hermann Pohle.

Kloesgen, W. and Żytkow, J. (eds.) (2002). *Handbook of Knowledge Discovery and Data Mining*, Oxford: Oxford University Press.

Lin, T.Y. and Cercone, N. (1997). *Rough Sets and Data Mining - Analysis of Imperfect Data*, Boston: Kluwer Academic Publishers.

Łukasiewicz, J. (1970). Die logischen Grundlagen der Wahrscheinlichkeitsrechnung (1913), *Jan Łukasiewicz - Selected Works*, Borkowski, L. (ed.), Amsterdam, Warsaw: Polish Scientific Publishers and North-Holland Publishing Company.

Orłowska, E. (1997). *Incomplete Information: Rough Set Analysis*, Physica-Verlag, Heidelberg.

Pal, S.K. and Skowron, A. (1999). *Rough Fuzzy Hybridization: A New Trend in Decision-Making*, Singapore: Springer-Verlag.

Pal, S.K., Polkowski, L. and Skowron, A. (eds.) (2004). *Rough--Neural Computing. Techniques for Computing with Words*, Berlin: Springer-Verlag.

Pawlak, Z. (1991). *Rough Sets: Theoretical Aspects of Reasoning about Data*, Dordrecht: Kluwer.

Pawlak, Z. (1982). Rough Sets, *Int. J. Computer Information Sci*., 11, 341—356.

Pawlak, Z. and Skowron, A. (1994). Rough membership functions, *Advances in the Dempster--Schafer Theory of Evidence*, Yager, R.R., Fedrizzi,M., and Kacprzyk, J. (eds.), New York: Wiley, 251—271.

Polkowski, L (2002). *Rough Sets: Mathematical Foundations.* Heidelberg: Physica-Verlag.

Polkowski, L. and Skowron, A. (eds.) (1998). *Rough Sets in Knowledge Discovery I & II,*

Heidelberg: Physica-Verlag.
Polkowski, L. and Skowron, A. (1996). Rough mereology: a new paradigm for approximate reasoning, *International Journal of Approximate Reasoning,* 15(4), 333—365.
Polkowski, L., Tsumoto, S. and Lin, T.Y. (eds.) (2000). *Rough Set Methods and Applications. New Developments in Knowledge Discovery in Information Systems*, Heidelberg: Physica-Verlag.
Russell, S.J. and Norvig, P. (2003). *Artificial Intelligence. A Modern Approach*, Prentice Hall, Saddle River, NJ.
Skowron, A. and Rauszer, C. (1992). The discernibility matrices and functions in information systems, *Intelligent Decision Support - Handbook of Applications and Advances of the Rough Sets Theory*, R. Slowiński (ed.), Dordrecht: Kluwer, 331-362.
Skowron, A. and Stepaniuk, J. (2001). Information granules: Towards foundations of granular computing, *International Journal of Intelligent Systems*, 16(1), 57-86.
Słowiński, R. (ed.) (1992). *Intelligent Decision Support - Handbook of Applications and Advances of the Rough Sets Theory*, Dordrecht: Kluwer Academic Publishers.
Zadeh, L.A. (1965). Fuzzy sets, *Information and Control*, 8, 338-353.
Zadeh, L.A. (2001). A new direction in AI: Toward a computational theory of perceptions, *AI Magazine*, 22(1), 73-84.
RSES Homepage: logic.mimuw.edu.pl
ROSETTA Homepage: rosetta.lcb.uu.se

Terms and Definitions

Information system is a pair (U,A) where U is the universe of objects and A is a set of attributes, i.e., functions on U with values in respective value sets V_a for $a \in A$.

Information about object x in a given information system (U,A) is defined by $Inf_A(x)=\{(a,a(x)): a \in A\}$.

Indiscernibility relation is defined as follows: objects x, y are indiscernible iff information about x is equal to (similar with) information about y. In the former case the indiscrnibility relation is an equivalence relation in the latter it is a similarity relation. Any object x defines an indiscernibility class (neighborhood) of objects indiscernible with this object. Also soft cases of indiscernibility relation are considered.

Discernibility relation is a binary relation on objects defined as follows: objects x,y are discernible iff information about x is discernible from information about y. In the simplest case objects x, y are discernible iff it is not true that they are indiscernible.

Rough set is a subset (concept) of the universe of objects U in an information system (U,A) that cannot be expressed (defined) as a union of indiscernibility classes; otherwise, the set is called **exact**.

Lower approximation of a set $X \subseteq U$ is the union of all indiscernibility classes contained in X, i.e., it is the greatest exact set contained in X. Instead of exact containment also containment to a degree is used.

Upper approximation of a set $X \subseteq U$ is the union of all indiscernibility classes that intersect X, i.e., it is the smallest exact set containing X.

Reduct is a minimal with respect to inclusion subset C of B preserving a given indiscernibility (discernibility) constraint, e.g., IND(C)=IND(B). Many different kinds of reducts with respect to different discernibility criteria have been investigated and used in searching for relevant patterns in data.

Functional dependence: for attribute sets C, D, we say that D depends functionally on C, in symbols

$C \rightarrow D$, in case $IND(C) \subseteq IND(D)$. Also non-exact (partial) functional dependencies to a degree are considered.

Decision system is a tuple (U,A,d), where (U,A) is an information system with the set A of condition attributes, and the decision (attribute) $d: U \rightarrow V_d$, where $d \notin A$. Informally, d is an attribute whose value is given by an external source (oracle, expert), in contradiction to conditional attributes in A whose values are determined by the user of the system.

Decision rule in (U,A,d) is any expression of the form $\wedge\{a=v_a : a \in A \text{ and } v_a \in V_a\} \rightarrow d=v$ where d is the decision attribute and v is a decision value. This decision rule is true in (U,A,d) if for any object satisfying its left hand side it also satisfies the right hand side, otherwise the decision rule is true to a degree measured by some coefficients such as confidence.

Bolean reasoning is based on construction for a given problem P of a corresponding Boolean function f_P with the following property: the solutions for the problem P can be decoded from prime implicants of the Boolean function f_P.