Rough Sets

Zdzisław Pawlak
Polish Academy of Sciences and Warsaw School of Information Technology, Poland

Lech Polkowski
Polish-Japanese Institute of Information Technology and University of Warmia and Mazury, Poland

Andrzej Skowron
Warsaw University, Poland

INTRODUCTION

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 (1965). Rough set theory (Pawlak, 1982) presents still another attempt at this problem. This theory has attracted the 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 the significance of data, and generating sets of decision rules from data. This approach is easy to understand and offers straightforward interpretation of obtained results, and 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 a 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 a nonempty boundary was made by Gottlob Frege in the beginning of 20th century.

BACKGROUND

In Table 1, we consider essential issues in rough sets in more detail.

ROUGH SETS AS A TOOL FOR REASONING ABOUT VAGUE CONCEPTS

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,
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 given 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 \( C \rightarrow D \) it 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 whose 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 \( [y]_B \subseteq X \) or \( [y]_B \cap X = \emptyset \). This implies that \( X \) has to be the union of a collection of \( B \)-indiscernibility classes, i.e., \( X = \bigcup \{ [x]_B : x \in X \} \). 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 \( B^{-}\langle Y \rangle = \{x \in X : [x]_B \subseteq X \} \) and \( B^{+}\langle Y \rangle = \{x \in X : [x]_B \cap X \neq \emptyset \} \), respectively. The set \( B^{-}\langle Y \rangle \) 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 Lukasiewicz (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 \( A' = (U, A, d) \), where \((U, A)\) is an information system with the set \( A \) of condition attributes, and the decision (attribute) \( d : U \rightarrow V_d \), where \( d \in A \). In case \( A \rightarrow d \) holds in \( A' \), we say that the decision system \( A' \) is deterministic and the dependency \( A \rightarrow d \) is \( A' \)-exact. Then, for each class \([x]_A\), there exists a unique decision \( d(x) \) throughout the class. Otherwise, the dependency \( A \rightarrow d \) in \( A' \) holds to a degree. A decision rule in \( A' \) 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).