Rough Sets - a New Paradigm of Imprecise Knowledge

Zdzisław Pawlak
Institute of Theoretical and Applied Informatics
Polish Academy of Sciences
ul. Bałtycka 5, 44 000 Gliwice, Poland
e-mail:zpw@ii.pw.edu.pl

1 Introduction

Theory of knowledge has been studied by philosophers and logicians for many years [2, 12, 44]. Besides, epistemology became a very important issue for researchers involved in AI and cognitive sciences, who contributed essentially to this domain [1, 3, 4, 5, 6, 7, 11, 25, 26, 30]. Nevertheless many issues pertinent to theory of knowledge, particularly in the context of AI, seem to be far from being fully understood. Especially the problem of imperfect knowledge requires due attention. Main flaws of imperfect knowledge are vagueness and uncertainty. Rough set theory, besides fuzzy set theory and the theory of evidence, contributed essentially to better understanding knowledge based systems, especially if vagueness and uncertainty are concerned. Many papers have been published on various aspects of knowledge and rough set theory [8, 9, 10, 13, 14, 15, 16, 17, 18, 27, 28, 35, 37, 40, 42, 43, 45, 46, 47, 48, 49, 54, 55, 59, 60, 61, 62, 64, 67, 69, 70, 71, 72, 73, 75, 77, 79, 80]. This paper concerns basic concepts of rough set theory with the emphasis on its relationship to knowledge. The references also include basic literature on rough sets and their applications [19, 20, 21, 22, 23, 24, 29, 34, 36, 37, 38, 39, 41, 51, 52, 53, 56, 57, 65, 66, 74, 76].

2 Basic concepts of rough set theory - approximations

Rough set theory is based on the assumption, widely shared in cognitive sciences, that the fundamental mechanism of reasoning is founded on the ability to classify elements of the universe of discourse. Classification means that small differences between elements are ignored and consequently those elements are indiscernible. Hence classification leads to clustering of elements of interest into granules, classes, clumps, groups, etc. of indiscernible (similar) objects. In rough set theory these granules, called elementary sets (concepts) form basic building blocks (concepts) of knowledge about the universe.

Every union of elementary concepts is referred to as a crisp or precise concept (set) otherwise a concept (set) is called rough, vague or imprecise. Thus rough concepts cannot be expressed in terms of elementary concepts. However, they can be expressed approximately by means of elementary concepts by employing the idea of the lower and the
**upper approximation** of a concept. The lower approximation of a concept is the union of all elementary concepts which are included in the concept, whereas the upper approximation is the union of all elementary concepts which have nonempty intersection with the concept. The difference between the lower and the upper approximation of the concept is its **boundary region**. Now it can easily be seen that a concept is rough if it has nonempty boundary region, i.e. its lower and upper approximation are nonidentical. Obviously, if the lower and the upper approximations of the concept are the same, i.e. its boundary region is empty – the concept is crisp.

Thus the basic flaw of imperfect knowledge, vagueness, can be remedied by replacing vague concepts by two precise concepts – its lower and upper approximation. These approximations are key ideas of rough set theory.

In what follows we shall formalize the above considerations and define more precisely basic concepts of rough set theory. Let us first discuss briefly the concept of a database. By a database we will understand a pair \( S = (U, A) \), where \( U \) and \( A \) are finite, nonempty sets called the **universe**, and a set **attributes** respectively. With every attribute \( a \in A \) we associate a set \( V_a \), of its **values**, called the **domain** of \( a \). Any subset \( B \) of \( A \) determines a binary relation \( I(B) \) on \( U \), which will be called an **indiscernibility relation**, and is defined as follows:

\[
(x, y) \in I(B) \text{ if and only if } a(x) = a(y) \text{ for every } a \in A, \text{ where } a(x) \text{ denotes the value of attribute } a \text{ for element } x.
\]

It can easily be seen that \( I(B) \) is an equivalence relation. The family of all equivalence classes of \( I(B) \), i.e. partition determined by \( B \), will be denoted by \( U/I(B) \), or simple \( U/B \); an equivalence class of \( I(B) \), i.e. block of the partition \( U/B \), containing \( x \) will be denoted by \( B(x) \).

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

There are several comments in order regarding the introduced definitions. The concept of a database used here is in fact a data table whose columns are labeled by attributes and rows – by elements of the universe. Such tables are also known as **information systems**. Thus database is simple a set of data about some elements of interest, e.g. patients in a hospital, cars, states of a process etc. An important point to note is that elements of the universe are described in the database by some features expressed by attribute values. It is rather straightforward to observe that this assumption leads to the indiscernibility relation, which results that some elements of the universe are clustered into **granules**, **classes**, **blocks**, **atoms**, etc. of indiscernible (similar) elements. These granules are treated as a whole, and they form the basic building blocks of our knowledge.

Now we defined two operations on sets:

\[
B_s(X) = \{x \in U : B(x) \subseteq X\},
B^*(X) = \{x \in U : B(x) \cap X \neq \emptyset\},
\]

which assign to every subset \( X \) of the universe \( U \) two sets \( B_s(X) \) and \( B^*(X) \) called the **\( B \)-lower** and the **\( B \)-upper approximation** of \( X \), respectively.

The set

\[
BN_B(X) = B^*(X) - B_s(X)
\]
will be called the \textit{B-boundary region} of \(X\).

If the boundary region of \(X\) is the empty set, i.e. \(BN_B(X) = \emptyset\), then the set \(X\) is \textit{crisp (exact)} with respect to \(B\); in the opposite case, i.e. if \(BN_B(X) \neq \emptyset\), the set \(X\) is \textit{rough (inexact)} with respect to \(B\).

It is important to emphasize that approximations are meant to substitute a pair of precise concepts for the imprecise concept.

Sets usually are defined by employing a membership function.

Rough sets can also be defined by using a \textit{rough membership function}, defined as

\[
\mu^B_X(x) = \frac{\text{card}(B(x) \cap X)}{\text{card}(B(x))}.
\]

Obviously

\[
\mu^B_X(x) \in [0,1].
\]

The value of the membership function \(\mu^B_X(x)\) is a kind of conditional probability, and can be interpreted as a degree of \textit{certainty} that \(x\) can be classified as \(X\) employing set of attributes \(B\).

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

\[
B_*(X) = \{x \in U : \mu^B_X(x) = 1\},
\]

\[
B^*(X) = \{x \in U : \mu^B_X(x) > 0\},
\]

\[
BN_B(X) = \{x \in U : 0 < \mu^B_X(x) < 1\}.
\]

It might seem that the rough membership function is identical with that used in fuzzy set theory, but this is not the case. For details the reader is referred to [39]. Thus we have two ways of defining rough sets: the first one uses approximations, whereas the second one employs the rough membership function. It is important to observe that these two approaches are not equivalent [39]. Approximations are in fact some topological operations on sets whereas the rough membership function, as mentioned before, is a kind of conditional probability. Approximations refer to \textit{vagueness} of a concept (set), but rough membership refers to \textit{uncertainty} whether some elements of the universe belong to a concept or not. Hence in rough set theory vagueness and uncertainty are clearly defined and understood and one can easily see the relationship between these concepts.

The rough membership function can be generalized as follows [41, 42]:

\[
\mu(X,Y) = \frac{\text{card}(X \cap Y)}{\text{card}X},
\]

where \(X,Y \subseteq U, X \neq \emptyset\).

Function \(\mu(X,Y)\) is an example of a \textit{rough inclusion} and expresses the degree to which \(X\) is included in \(Y\). Obviously, if \(\mu(X,Y) = 1\), then \(X \subseteq Y\).

If \(X\) is included to a degree \(k\) we will write \(X \subseteq_k Y\).

The rough inclusion function can be interpreted as a generalization of the mereological relation ”part of”, and reads as ”part to a degree” [41, 42]. We will use this construction in Section 5.
3 Dependency of Attributes

Instead of using approximations of sets we can use the concept of dependency of attributes.

Intuitively, a set of attributes \( D \) (called decision attributes) depends totally on a set of attributes \( C \) (called condition attributes), denoted \( C \Rightarrow D \), if all values of attributes from \( D \) are uniquely determined by values of attributes from \( C \). In other words, \( D \) depends totally on \( C \), if there exists a functional dependency between values of \( D \) and \( C \).

Formally, dependency can be defined in the following way.

Let \( D \) and \( C \) be subsets of \( A \). We say that \( D \) depends totally on \( C \), if \( I(C) \subseteq I(D) \). That means that the partition generated by \( C \) is finer than the partition generated by \( D \).

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

Formally, the above idea can be formulated as follows.

Let \( D \) and \( C \) be subsets of \( A \). We say that \( D \) depends to a degree \( k \), \( 0 \leq k \leq 1 \), on \( C \), denoted \( C \Rightarrow_k D \), if

\[
k = \gamma(C, D) = \frac{\text{card}(\text{POS}_C(D))}{\text{card}(U)} = \frac{\sum_{x \in U/D} \text{card}(C_*(X))}{\text{card}(U)},
\]

where

\[
\text{POS}_C(D) = \bigcup_{X \in U/I(D)} C_*(X).
\]

In our approach partial dependencies and approximations are used to express vague-ness. If we want to deal with the global picture of vague patterns in a database the use of dependencies is in order, because they show possible "cause-effect" relations, or approximate dependencies, occurring in the database. However, if we are interested in local properties of a database and want to know how some concepts can be expressed in terms of elementary concepts, approximations are the answer to this problem.

4 Reduction of Attributes

Another important issue in our approach is data reduction.

This concept can be formulated as follows. Let \( C \Rightarrow_k D \). A minimal subset \( C' \) of \( C \), such that \( \gamma(C, D) = \gamma(C', D) \) is called a reduct of \( C \).

Thus a reduct is a set of condition attributes that preserves the degree of dependency. It means that a reduct is a minimal subset of condition attributes that enables the same decisions as the whole set of condition attributes.

Obviously a set of condition attributes may have more than one reduct. Intersection of all reducts is called the core. The core is the set of attributes that cannot be eliminated from the information table without changing its dependencies and approximations.

In other words, attribute reduction shows how data can be reduced from a database without affecting its basic properties. This is the fundamental issue in rough set theory. Many effective methods of attribute reduction have been proposed and implemented. Nevertheless effective methods of reducts computation are still badly needed, particularly when very large databases are concerned.
5 Dependencies, Decision Rules and Knowledge Base

Each dependency $C \Rightarrow_k D$ in a database induce a set of decision rules of the form "if... then...", called a knowledge base. In other words every dependency $C \Rightarrow_k D$ can be represented by a set of decision rules:

\[
\begin{align*}
&\text{if } C_1 \text{ then } D_1 \\
&\text{if } C_2 \text{ then } D_2 \\
&\ldots \\
&\text{if } C_n \text{ then } D_n
\end{align*}
\]

where $C_i$ and $D_i$ are sets of conditions and decisions, respectively.

Each decision rule corresponds to a row in a database and represents decisions that should be made when conditions specified by the rule are satisfied. Thus knowledge base is understood here as a set of decision rules. This view is widely shared in the AI community.

Decision rules are implications, therefore with every database $S = (U, A)$ we associate a formal language. The language is defined in the standard way and we assume that the reader is familiar with the construction.

Given $x \in U$ and $B \subseteq A$ by $\Phi^B_x = \bigwedge_{a \in B}(a, v)$ we mean a formula such that $a(x) = v$ and $v \in V_a$.

Every dependency $C \Rightarrow_k D$ determines a set of decision rules (knowledge base)

\[
\{\Phi^C_x \rightarrow \Phi^D_x\}_{x \in U}.
\]

We say that a decision rule $\Phi^C_x \rightarrow \Phi^D_x$ is true in $S$, if $|\Phi^C_x|_S \subseteq |\Phi^D_x|_S$, where $|\Phi^C_x|_S$ denotes the meaning of $\Phi^C_x$ in $S$, i.e. the set of all $y \in U$ that satisfy $\Phi^C_x$ in $S$.

Let $C_S(x) = |\Phi^C_x|_S$. Hence the decision rule $\Phi^C_x \rightarrow \Phi^D_x$ is true in $S$ if $C_S(x) \subseteq D_S(x)$.

A decision rule $\Phi^C_x \rightarrow \Phi^D_x$ is true to a degree $l$ in $S$, if $l = \mu(C_S(x), D_S(x)) > 0$, i.e. $C_S(x) \subseteq l D_S(x)$.

Rough inclusion in this case boils down to rough membership function. As a consequence rough membership can be interpreted as a generalized truth value.

The degree of truth of a decision rule can also be interpreted as a certainty factor of the rule.

Let us observe that the rough membership can be interpreted both as conditional probability and at the same time as partial truth value.

The above considerations lead to a inference rule, called the rough modus ponens and is defined as below:

\[
\frac{\pi(\Phi^C_x); \mu(\Phi^C_x, \Psi^D_x)}{\pi(\Psi^C_x)},
\]

where

\[
\pi(\Phi^C_x) = \frac{\text{card}(|\Phi^C_x|_S)}{\text{card} U},
\]

\[
\mu(\Phi^C_x, \Psi^D_x) = \frac{\text{card}(|\Phi^C_x \land \Psi^D_x|_S)}{\text{card} |\Phi^C_x|_S}
\]
and
\[ \pi(\Psi^D_x) = \sum_{y \in D(x)} (\pi(\Phi^C_y) \cdot \mu(\Phi^C_y, \Psi^D_y)). \]

The number \( \pi(\Phi^C_x) \) can be interpreted as the probability, that \( x \) has the property \( \Phi^C_x \), and the number \( \mu(\Phi^C_x, \Psi^D_x) \) – as certainty factor of the decision rule \( \Phi^C_x \rightarrow \Psi^D_x \).

Hence the inference rule, the rough modus ponens, enables us to calculate the probability of conclusion \( \Psi^D_x \) as a function of the probability of the premise \( \Phi^C_x \) and the certainty factor \( \mu(\Phi^D_x, \Psi^D_x) \) of the decision rule \( \Phi^C_x \rightarrow \Psi^D_x \).

6 Conclusions

Because knowledge base can be treated as a set of decision rules (implications) basic concepts of rough set theory can be expressed not only in algebraic terms but also in logical framework. Many logical systems (rough logics) based on these ideas have been proposed and investigated, but we have not discussed these issues here. We would only like to stress that in view of the above considerations rough set theory leads to another approach to reasoning about knowledge: reasoning in rough set theory can be based on rough modus ponens. In contrast to modus ponens – which allows to draw true conclusions from true premises by means of true implications – rough modus ponens enables to evaluate the probability of conclusions on the basic of probabilities of premises and the certainty factor of decision rules involved.

References


[43] L. Polkowski, A. Skowron (eds.): Rough Sets in Knowledge Discovery. Springer Verlag (in print)


