21. Foundations of Rough Sets

Andrzej Skowron, Andrzej Jankowski, Roman W. Swiniarski (deceased)

The rough set (RS) approach was proposed by Pawlak as a tool to deal with imperfect knowledge. Over the years the approach has attracted attention of many researchers and practitioners all over the world, who have contributed essentially to its development and applications. This chapter discusses the RS foundations from rudiments to challenges.

21.1 Rough Sets: Comments on Development

The rough set (RS) approach was proposed by *Zdzisław Pawlak* in 1982 [21.1, 2] as a tool for dealing with imperfect knowledge, in particular with vague concepts. Many applications of methods based on rough set theory alone or in combination with other approaches have been developed. This chapter discusses the RS foundations from rudiments to challenges.

In the development of rough set theory and applications, one can distinguish three main stages. While the first period was based on the assumption that objects are perceived by means of partial information represented by attributes, in the second period it was assumed that information about the approximated concepts is also partial. Approximation spaces and searching strategies for relevant approximation spaces were recognized as the basic tools for rough sets. Important achievements both in theory and applications were obtained. Nowadays, a new period for rough sets is emerging, which is also briefly characterized in this chapter.

The rough set approach seems to be of fundamental importance in artificial intelligence AI and cognitive sciences, especially in machine learning, data mining,

knowledge discovery from databases, pattern recognition, decision support systems, expert systems, intelligent systems, multiagent systems, adaptive systems, autonomous systems, inductive reasoning, commonsense reasoning, adaptive judgment, conflict analysis.

Rough sets have established relationships with many other approaches such as fuzzy set theory, granular computing (GC), evidence theory, formal concept analysis, (approximate) Boolean reasoning, multicriteria decision analysis, statistical methods, decision theory, and matroids. Despite the overlap with many other theories rough set theory may be considered as an independent discipline in its own right. There are reports on many hybrid methods obtained by combining rough sets with other approaches such as soft computing (fuzzy sets, neural networks, genetic algorithms), statistics, natural computing, mereology, principal component analysis, singular value decomposition, or support vector machines.

The main advantage of rough set theory in data analysis is that it does not necessarily need any preliminary or additional information about data like probability distributions in statistics, basic probability assignments

in evidence theory, a grade of membership, or the value of possibility in fuzzy set theory.

One can observe the following advantages about the rough set approach:

i)   Introduction of efficient algorithms for finding hidden patterns in data.
ii) Determination of optimal sets of data (data reduction); evaluation of the significance of data.
iii) Generation of sets of decision rules from data.
iv) Easy-to-understand formulation.
v)   Straightforward interpretation of results obtained.
vi) Suitability of many of its algorithms for parallel processing.

Due to space limitations, many important research topics in rough set theory such as various logics related to rough sets and many advanced algebraic properties of rough sets are only mentioned briefly in this chapter.

From the same reason, we herein restrict the references on rough sets to the basic papers by *Zdzisław Pawlak* (such as [21.1, 2]), some survey papers [21.3–5], and some books including long lists of references to papers on rough sets. The basic ideas of rough set theory and its extensions as well as many interesting applications can be found in a number of books, issues

of *Transactions on Rough Sets*, special issues of other journals, numerous proceedings of international conferences, and tutorials [21.3, 6, 7]. The reader is referred to the cited books and papers, references therein, as well as to web pages [21.8, 9].

The chapter is structured as follows. In Sect. 21.2 we discuss some basic issues related to vagueness and vague concepts. The rough set philosophy is outlined in Sect. 21.3. The basic concepts for rough sets such as indiscernibility and approximation are presented in Sect. 21.4. Decision systems and rules are covered in Sect. 21.5. The basic information about dependencies is included in Sect. 21.6. Attribute reduction belonging to one of the basic problems of rough sets is discussed in Sect. 21.7. Rough membership function as a tool for measuring degrees of inclusion of sets is presented in Sect. 21.8. The role of discernibility and Boolean reasoning for solving problems related to rough sets is briefly explained in Sect. 21.9. In Sect. 21.10 a short discussion on rough sets and induction is included. Several generalizations of the approach proposed by Pawlak are discussed in Sect. 21.11. In this section some emerging research directions related to rough sets are also outlined. In Sect. 21.12 some comments about logics based on rough sets are included. The role of adaptive judgment is emphasized.

21.2 Vague Concepts

Mathematics requires that all mathematical notions (including sets) must be exact, otherwise precise reasoning would be impossible. However, philosophers [21.10], and recently computer scientists as well as other researchers, have become interested in *vague* (imprecise) concepts. Moreover, in the twentieth century one can observe the drift paradigms in modern science from dealing with precise concepts to vague concepts, especially in the case of complex systems (e.g., in economy, biology, psychology, sociology, and quantum mechanics).

In classical set theory, a set is uniquely determined by its elements. In other words, this means that every element must be uniquely classified as belonging to the set or not. That is to say the notion of a set is a *crisp* (precise) one. For example, the set of odd numbers is crisp because every integer is either odd or even.

In contrast to odd numbers, the notion of a beautiful painting is vague, because we are unable to classify uniquely all paintings into two classes: *beautiful* and

not beautiful. With some paintings it cannot be decided whether they are beautiful or not and thus they remain in the doubtful area. Thus, *beauty* is not a precise but a vague concept.

Almost all concepts that we use in natural language are vague. Therefore, common sense reasoning based on natural language must be based on vague concepts and not on classical logic. An interesting discussion of this issue can be found in [21.11]. The idea of vagueness can be traced back to the ancient Greek philosopher Eubulides of Megara (ca. 400 BC) who first formulated the so-called *sorites* (heap) and *falakros* (bald man) paradoxes [21.10]. There is a huge literature on issues related to vagueness and vague concepts in philosophy [21.10].

Vagueness is often associated with the boundary region approach (i. e., existence of objects which cannot be uniquely classified relative to a set or its complement), which was first formulated in 1893 by the father of modern logic, the German logician, *Gottlob Frege* (1848–1925) ([21.12]). According to Frege the concept

must have a sharp boundary. To the concept without a sharp boundary there would correspond an area that would not have any sharp boundary line all around. This means that mathematics must use crisp, not vague concepts, otherwise it would be impossible to reason precisely.

One should also note that vagueness also relates to insufficient specificity, as the result of a lack of feasible searching methods for sets of features adequately describing concepts. A discussion on vague (imprecise) concepts in philosophy includes their following characteristic features [21.10]: (i) the presence of borderline cases, (ii) boundary regions of vague concepts are not crisp, (iii) vague concepts are susceptible to sorites paradoxes. In the sequel we discuss the first two issues in the RS framework. The reader can find a discussion on the application of the RS approach to the third item in [21.11].

21.3 Rough Set Philosophy

Rough set philosophy is founded on the assumption that with every object of the universe of discourse we associate some information (data, knowledge). For example, if objects are patients suffering from a certain disease, symptoms of the disease form information about the 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. This understanding of indiscernibility is related to the idea of *Gottfried Wilhelm Leibniz* that objects are indiscernible if and only if all available functionals take identical values on them (Leibniz's law of indiscernibility: the identity of indiscernibles) [21.13]. However, in the rough set approach indiscernibility is defined relative to a given set of functionals (attributes).

Any set of all indiscernible (similar) objects is called an elementary set and forms a basic granule (atom) of knowledge about the universe. Any union of some elementary sets is referred to as a *crisp* (precise) set. If a set is not crisp, then it is called *rough* (imprecise, vague). Consequently, each rough set has *borderline cases* (*boundary-line*), i. e., objects which cannot be classified with certainty as members of either the set or its complement. Obviously, crisp sets have no borderline elements at all. This means that borderline cases cannot be properly classified by employing available knowledge.

Thus, the assumption that objects can be *seen* only through the information available about them leads to the view that knowledge has granular structure. Due to the granularity of knowledge, some objects of interest cannot be discerned and appear as the same (or similar). As a consequence, vague concepts in contrast to precise concepts, cannot be characterized in terms of information about their elements. Therefore, in the proposed approach, we assume that any vague concept is 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 definitely 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. Approximation operations are the basic operations in rough set theory. Properties of the boundary region (expressed, e.g., by the rough membership function) are important in the rough set methods.

Hence, rough set theory expresses vagueness not by means of membership, but by employing a boundary region of a set. If the boundary region of a set is empty it means that the set is crisp, otherwise the set is rough (inexact). A nonempty boundary region of a set means that our knowledge about the set is not sufficient to define the set precisely.

Rough set theory it is not an alternative to classical set theory but it is embedded in it. Rough set theory can be viewed as a specific implementation of Frege's idea of vagueness, i. e., imprecision in this approach is expressed by a boundary region of a set.

21.4 Indiscernibility and Approximation

The starting point of rough set theory is the indiscernibility relation, which is generated by information about objects of interest (Sect. 21.1). The indiscernibility relation expresses the fact that due to a lack of information (or knowledge) we are unable to discern some objects by employing available information (or knowledge).

This means that, in general, we are unable to deal with each particular object but we have to consider granules (clusters) of indiscernible objects as a fundamental basis for our theory.

From a practical point of view, it is better to define basic concepts of this theory in terms of data. Therefore, we will start our considerations from a data set called an *information system*. An information system can be represented by a data table containing rows labeled by objects of interest and columns labeled by attributes and entries of the table are attribute values. For example, a data table can describe a set of patients in a hospital. The patients can be characterized by some attributes, like *age, sex, blood pressure, body temperature*, etc. With every attribute a set of its values is associated, e.g., values of the attribute *age* can be *young, middle*, and *old*. Attribute values can also be numerical. In data analysis the basic problem that we are interested in is to find patterns in data, i. e., to find a relationship between some set of attributes, e.g., we might be interested whether *blood pressure* depends on *age and sex*.

More formally, suppose we are given a pair $\mathcal{A} = (U, A)$ of nonempty, finite sets U and A, where U is the *universe* of *objects*, and an $A - a$ set consisting of *attributes*, i. e., functions $a : U \longrightarrow V_a$, where V_a is the set of values of attribute a, called the *domain* of a. The pair $\mathcal{A} = (U, A)$ is called an *information system*. Any information system can be represented by a data table with rows labeled by objects and columns labeled by attributes. Any pair (x, a), where $x \in U$ and $a \in A$ defines the table entry consisting of the value $a(x)$. Note that in statistics or machine learning such a data table is called a *sample* [21.14].

Any subset B of A determines a binary relation $IND(B)$ on U, called an *indiscernibility relation*, defined by

$$x IND(B) y \text{ if and only if} \\ a(x) = a(y) \text{ for every } a \in B , \tag{21.1}$$

where $a(x)$ denotes the value of attribute a for the object x.

Obviously, $IND(B)$ is an equivalence relation. The family of all equivalence classes of $IND(B)$, i. e., the partition determined by B, will be denoted by $U/IND(B)$, or simply U/B; the equivalence class of $IND(B)$, i. e., the block of the partition U/B, containing x will be denoted by $B(x)$ (other notation used: $[x]_B$ or $[x]_{IND(B)}$). Thus in view of the data we are unable, in general, to observe individual objects but we are forced to reason only about the accessible granules of knowledge.

If $(x, y) \in IND(B)$ we will say that x and y are *B-indiscernible*. Equivalence classes of the relation $IND(B)$ (or blocks of the partition U/B) are referred to as *B-elementary sets* or *B-elementary granules*. In the rough set approach the elementary sets are the basic building blocks (concepts) of our knowledge about reality. The unions of *B-elementary sets* are called *B-definable sets*. Let us note that in applications we consider only some subsets of the family of definable sets, e.g., defined by conjunction of descriptors only. This is due to the computational complexity of the searching problem for relevant definable sets in the whole family of definable sets.

For $B \subseteq A$, we denote by $Inf_B(x)$ the *B*-signature of $x \in U$, i. e., the set $\{(a, a(x)) : a \in B\}$. Let $INF(B) = \{Inf_B(x) : x \in U\}$. Then for any objects $x, y \in U$ the following equivalence holds: $x IND(B) y$ if and only if $Inf_B(x) = Inf_B(y)$.

The indiscernibility relation is used to define the approximations of concepts. We define the following two operations on sets $X \subseteq U$

$$B_*(X) = \{x \in U : B(x) \subseteq X\} , \tag{21.2}$$
$$B^*(X) = \{x \in U : B(x) \cap X \neq \emptyset\} , \tag{21.3}$$

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

$$BN_B(X) = B^*(X) - B_*(X) , \tag{21.4}$$

will be referred to as the *B-boundary region* of X.

From the definition we obtain the following interpretation: (i) the *lower approximation* of a set X with respect to B is the set of all objects, which can for *certain* be classified to X using B (are *certainly* in X in view of B), (ii) the *upper approximation* of a set X with respect to B is the set of all objects which can *possibly* be classified to X using B (are *possibly* in X in view of B), (iii) the *boundary region* of a set X with respect to B is the set of all objects, which can be classified neither to X nor to not-X using B.

Due to the granularity of knowledge, rough sets cannot be characterized by using available knowledge. The definition of approximations is clearly depicted in Fig. 21.1.

The approximations have the following properties

$$B_*(X) \subseteq X \subseteq B^*(X) \,,$$
$$B_*(\emptyset) = B^*(\emptyset) = \emptyset \,, B_*(U) = B^*(U) = U \,,$$
$$B^*(X \cup Y) = B^*(X) \cup B^*(Y) \,,$$
$$B_*(X \cap Y) = B_*(X) \cap B_*(Y) \,,$$
$$X \subseteq Y \text{ implies } B_*(X) \subseteq B_*(Y)$$
$$\text{and } B^*(X) \subseteq B^*(Y) \,,$$
$$B_*(X \cup Y) \supseteq B_*(X) \cup B_*(Y) \,,$$
$$B^*(X \cap Y) \subseteq B^*(X) \cap B^*(Y) \,,$$
$$B_*(-X) = -B^*(X) \,,$$
$$B^*(-X) = -B_*(X) \,,$$
$$B_*(B_*(X)) = B^*(B_*(X)) = B_*(X) \,,$$
$$B^*(B^*(X)) = B_*(B^*(X)) = B^*(X) \,. \tag{21.5}$$

Let us note that the inclusions (for union and intersection) in (21.5) cannot, in general, be substituted by the equalities. This has some important algorithmic and logical consequences.

Now we are ready to give the definition of rough sets. If the boundary region of X is the empty set, i.e., $\mathrm{BN}_B(X) = \emptyset$, then the set X is *crisp* (*exact*) with respect to B; in the opposite case, i.e., if $\mathrm{BN}_B(X) \neq \emptyset$, the set X is referred to as *rough* (*inexact*) with respect to B. Thus any rough set, in contrast to a crisp set, has a nonempty boundary region. This is the idea of vagueness proposed by Frege.
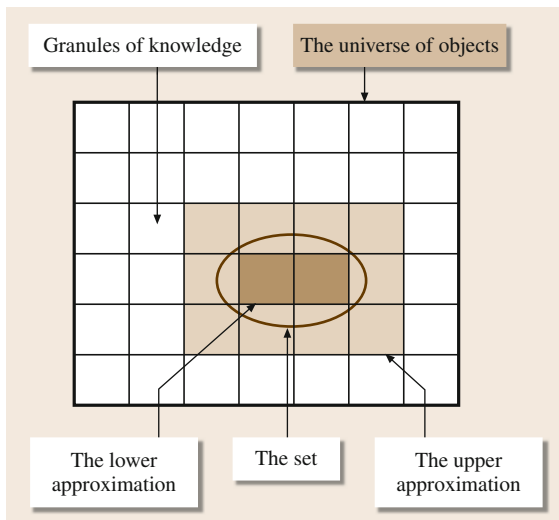

Fig. 21.1 A rough set

Let us observe that the definition of rough sets refers to data (knowledge), and is *subjective*, in contrast to the definition of classical sets, which is in some sense an *objective* one.

A rough set can also be characterized numerically by the following coefficient

$$\alpha_B(X) = \frac{\mathrm{card}(B_*(X))}{\mathrm{card}(B^*(X))} \,, \tag{21.6}$$

called the *accuracy of approximation*, where $X \neq \emptyset$ and $\mathrm{card}(X)$ denotes the cardinality of X. Obviously $0 \leq \alpha_B(X) \leq 1$. If $\alpha_B(X) = 1$ then X is *crisp* with respect to B (X is *precise* with respect to B), and otherwise, if $\alpha_B(X) < 1$ then X is *rough* with respect to B (X is *vague* with respect to B). The accuracy of approximation can be used to measure the quality of approximation of decision classes on the universe U. One can use another measure of accuracy defined by $1 - \alpha_B(X)$ or by

$$1 - \frac{\mathrm{card}(\mathrm{BN}_B(X))}{\mathrm{card}(U)} \,.$$

Some other measures of approximation accuracy are also used, e.g., based on entropy or some more specific properties of boundary regions. The choice of a relevant accuracy of approximation depends on a particular data set. Observe that the accuracy of approximation of X can be tuned by B. Another approach to the accuracy of approximation can be based on the variable precision rough set model (VPRSM).

In [21.10], it is stressed that boundaries of vague concepts are not crisp. In the definition presented in this chapter, the notion of boundary region is defined as a crisp set $\mathrm{BN}_B(X)$. However, let us observe that this definition is relative to the subjective knowledge expressed by attributes from B. Different sources of information may use different sets of attributes for concept approximation. Hence, the boundary region can change when we consider these different views. Another reason for boundary change may be related to incomplete information about concepts. They are known only on samples of objects. Hence, when new objects appear again the boundary region may change. From the discussion in the literature it follows that vague concepts cannot be approximated with satisfactory quality by *static* constructs such as induced membership inclusion functions, approximations, or models derived, e.g., from a sample. An understanding of vague concepts can be only realized in a process in which the induced models adaptively

match the concepts in dynamically changing environments. This conclusion seems to have important consequences for the further development of rough set theory in combination with fuzzy sets and other soft computing paradigms for adaptive approximate reasoning.

21.5 Decision Systems and Decision Rules

In this section, we discuss the decision rules (constructed over a selected set B of features or a family of sets of features), which are used in inducing classification algorithms (classifiers), making it possible to classify unseen objects to decision classes. Parameters which are tuned in searching for a classifier with high quality are its description size (defined, e.g., by used decision rules) and its quality of classification (measured, e.g., by the number of misclassified objects on a given set of objects). By selecting a proper balance between the accuracy of classification and the description size one can search for classifier with a high quality of classification also on testing objects. This approach is based on the minimum description length principle (MDL) [21.15].

In an information system $\mathcal{A} = (U, A)$ we sometimes distinguish a partition of A into two disjoint classes $C, D \subseteq A$ of attributes, called *condition* and *decision* (*action*) attributes, respectively. The tuple $\mathcal{A} = (U, C, D)$ is called a *decision system* (or a *decision table*).

Let $V = \bigcup \{V_a : a \in C\} \cup \bigcup \{V_d \mid d \in D\}$. Atomic formulae over $B \subseteq C \cup D$ and V are expressions $a = v$ called *descriptors (selectors) over B and V*, where $a \in B$ and $v \in V_a$. The set of formulae over B and V, denoted by $\mathcal{F}(B, V)$, is the least set containing all atomic formulae over B and V and closed under the propositional connectives $\wedge$ (conjunction), $\vee$ (disjunction) and $\neg$ (negation). By $\|\varphi\|_{\mathcal{A}}$ we denote the meaning of $\varphi \in \mathcal{F}(B, V)$ in the decision system $\mathcal{A}$, which is the set of all objects in U with the property φ. These sets are defined by

$$\|a = v\|_{\mathcal{A}} = \{x \in U \mid a(x) = v\} ,$$
$$\|\varphi \wedge \varphi'\|_{\mathcal{A}} = \|\varphi\|_{\mathcal{A}} \cap \|\varphi'\|_{\mathcal{A}} ;$$
$$\|\varphi \vee \varphi'\|_{\mathcal{A}} = \|\varphi\|_{\mathcal{A}} \cup \|\varphi'\|_{\mathcal{A}} ;$$
$$\|\neg\varphi\|_{\mathcal{A}} = U - \|\varphi\|_{\mathcal{A}} .$$

The formulae from $\mathcal{F}(C, V)$, $\mathcal{F}(D, V)$ are called *condition formulae of* $\mathcal{A}$ and *decision formulae of* $\mathcal{A}$, respectively.

Any object $x \in U$ belongs to the *decision class* $\| \bigwedge_{d \in D} d = d(x) \|_{\mathcal{A}}$ of $\mathcal{A}$. All decision classes of $\mathcal{A}$ create a partition U/D of the universe U.

A *decision rule* for $\mathcal{A}$ is any expression of the form $\varphi \Rightarrow \psi$, where $\varphi \in \mathcal{F}(C, V)$, $\psi \in \mathcal{F}(D, V)$, and $\|\varphi\|_{\mathcal{A}} \neq \emptyset$. Formulae φ and ψ are referred to as the *predecessor* and the *successor* of decision rule $\varphi \Rightarrow \psi$. Decision rules are often called *IF ... THEN ...* rules. Such rules are used in machine learning.

Decision rule $\varphi \Rightarrow \psi$ is *true* in $\mathcal{A}$ if and only if $\|\varphi\|_{\mathcal{A}} \subseteq \|\psi\|_{\mathcal{A}}$. Otherwise, one can measure its *truth degree* by introducing some inclusion measure of $\|\varphi\|_{\mathcal{A}}$ in $\|\psi\|_{\mathcal{A}}$. Let us denote by $\mathrm{card}_{\mathcal{A}}(\varphi)$ (or $\mathrm{card}(\varphi)$, if this does not lead to confusion) the number of objects from U that satisfies formula φ, i. e., the cardinality of $\|\varphi\|_{\mathcal{A}}$. According to *Łukasiewicz* [21.16], one can assign to formula φ the value

$$\frac{\mathrm{card}(\varphi)}{\mathrm{card}(U)} ,$$

and to the implication $\varphi \Rightarrow \psi$ the fractional value

$$\frac{\mathrm{card}(\varphi \wedge \psi)}{\mathrm{card}(\varphi)} ,$$

under the assumption that $\|\varphi\|_{\mathcal{A}} \neq \emptyset$. The fractional part proposed by Łukasiewicz was adapted much later by machine learning and data mining community, e.g., in the definitions of the accuracy of decision rules or confidence of association rules.

For any decision system $\mathcal{A} = (U, C, D)$ one can consider a *generalized decision* function $\delta_A : U \longrightarrow \mathrm{POW}(\mathrm{INF}(D))$, where for any $x \in U$, $\delta_A(x)$ is the set of all D-signatures of objects from U which are C-indiscernible with x, $A = C \cup D$ and $\mathrm{POW}(\mathrm{INF}(D))$ is the powerset of the set $\mathrm{INF}(D)$ of all possible decision signatures.

The decision system $\mathcal{A}$ is called *consistent (deterministic)*, if $\mathrm{card}(\delta_A(x)) = 1$, for any $x \in U$. Otherwise $\mathcal{A}$ is said to be *inconsistent (nondeterministic)*. Hence, a decision system is inconsistent if it consists

of some objects with different decisions but that are indiscernible with respect to condition attributes. Any set consisting of all objects with the same generalized decision value is called a *generalized decision class*.

Now, one can consider certain (possible) rules for decision classes defined by the lower (upper) approximations of such generalized decision classes of $\mathcal{A}$. This approach can be extended by using the relationships of rough sets with the evidence theory (Dempster–Shafer theory) by considering rules relative to decision classes defined by the lower approximations of unions of decision classes of $\mathcal{A}$.

Numerous methods have been developed for the generation of different types of decision rules, and the reader is referred to the literature on rough sets for details. Usually, one is searching for decision rules that are (semi) optimal with respect to some optimization criteria describing the quality of decision rules in concept approximations.

21.6 Dependencies

Another important issue in data analysis is discovering dependencies between attributes in a given decision system $\mathcal{A} = (U, C, D)$. Intuitively, a set of attributes D depends totally on a set of attributes C, denoted $C \Rightarrow D$, if the values of attributes from C uniquely determine the values of attributes from D. In other words, D depends totally on C, if there exists a functional dependency between values of C and D.

D can depend partially on C. Formally such a dependency can be defined in the following way. We will say that D *depends on* C to a *degree* k ($0 \leq k \leq 1$), denoted by $C \Rightarrow_k D$, if

$$k = \gamma(C, D) = \frac{\text{card}(\text{POS}_C(D))}{\text{card}(U)}, \tag{21.7}$$

where

$$\text{POS}_C(D) = \bigcup_{X \in U/D} C_*(X), \tag{21.8}$$

which is 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.

If $k = 1$, we say that D *depends totally* on C, and if $k < 1$, we say that D *depends partially* (to *degree k*) on C. If $k = 0$, then the *positive region* of the partition U/D with respect to C is empty.

The coefficient k expresses the ratio of all elements of the universe, which can be properly classified to blocks of the partition U/D, employing attributes C and is called the *degree of the dependency*.

It can be easily seen that if D depends totally on C, then $\text{IND}(C) \subseteq \text{IND}(D)$. This means that the partition generated by C is finer than the partition generated by D.

Summing up: D is *totally* (*partially*) dependent on C, if *all* (*some*) elements of the universe U can be uniquely classified to blocks of the partition U/D, employing C. Observe that (21.7) defines only one of the possible measures of dependency between attributes. Note that one can consider dependencies between arbitrary subsets of attributes in the same way. One also can compare the dependency discussed in this section with dependencies considered in databases.

21.7 Reduction of Attributes

We often face the question as to whether we can remove some data from a data table and still preserve its basic properties, that is – whether a table contains some superfluous data. 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

$$\gamma(C, D) = \gamma(C', D). \tag{21.9}$$

The intersection of all D-reducts is called a *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, since none of its elements can be removed without affecting the classification power of attributes. Certainly, the geometry of reducts can be more comlex. For example, the core can be empty but there can exist a partition of reducts into a few sets with nonempty intersection.

Many other kinds of reducts and their approximations have been discussed in the literature. They are defined relative to different quality measures. For example, if one changes the condition (21.9) to $\partial_A(x) = \partial_B(x)$, (where $A = C \cup D$ and $B = C' \cup D$), then the defined reducts preserve the generalized decision. Other kinds of reducts preserve, e.g., (i) the distance between attribute value vectors for any two objects, if this distance is greater than a given threshold, (ii) the distance between entropy distributions between any two objects, if this distance exceeds a given. Yet another kind of reducts is defined by the so-called reducts relative to object used for the generation of decision rules.

Reducts are used for building data models. Choosing a particular reduct or a set of reducts has an impact on the model size as well as on its quality in describing a given data set. The model size together with the model quality are two basic components tuned in selecting relevant data models. This is known as the minimum length principle. Selection of relevant kinds of reducts is an important step in building data models. It turns out that the different kinds of reducts can be efficiently computed using heuristics based, e.g., on the Boolean reasoning approach.

Let us note that analogously to the information flow [21.17] one can consider different theories over information or decision systems representing different views on knowledge encoded in the systems. In particular, this approach was used for inducing concurrent models from data tables. For more details the reader is referred to the books cited at the beginning of the chapter.

21.8 Rough Membership

Let us observe that rough sets can be also defined employing the rough membership function (21.10) instead of approximation. That is, consider

$$\mu_X^B : U \to [0, 1] \,,$$

defined by

$$\mu_X^B(x) = \frac{\text{card}(B(x) \cap X)}{\text{card}(B(x))} \,, \tag{21.10}$$

where $x \in X \subseteq U$. The value $\mu_X^B(x)$ can be interpreted as the degree that x belongs to X in view of knowledge about x expressed by B or the degree to which the elementary granule $B(x)$ is included in the set X. This means that the definition reflects a subjective knowledge about elements of the universe, in contrast to the classical definition of a set related to objective knowledge.

Rough membership function can also be interpreted as the conditional probability that x belongs to X given B. One may refer to Bayes' theorem as the origin of this function. This interpretation was used by several researchers in the rough set community.

One can observe that the rough membership function has the following properties:

1) $\mu_X^B(x) = 1$ iff $x \in B_*(X)$,
2) $\mu_X^B(x) = 0$ iff $x \in U - B^*(X)$,
3) $0 < \mu_X^B(x) < 1$ iff $x \in \text{BN}_B(X)$,
4) $\mu_{U-X}^B(x) = 1 - \mu_X^B(x)$ for any $x \in U$,
5) $\mu_{X \cup Y}^B(x) \geq \max(\mu_X^B(x),\ \mu_Y^B(x))$ for any $x \in U$,
6) $\mu_{X \cap Y}^B(x) \leq \min(\mu_X^B(x),\ \mu_Y^B(x))$ for any $x \in U$.

From the properties it follows that the rough membership differs essentially from the fuzzy membership [21.18], for properties 5) and 6) show that the membership for union and intersection of sets, in general, cannot be computed – as in the case of fuzzy sets – from their constituents' membership. Thus formally rough membership is different from fuzzy membership. Moreover, the rough membership function depends on available knowledge (represented by attributes from B). Besides, the rough membership function, in contrast to the fuzzy membership function, has a probabilistic flavor.

Let us also mention that rough set theory, in contrast to fuzzy set theory, clearly distinguishes two very important concepts, vagueness and uncertainty, very often confused in the AI literature. Vagueness is the property of concepts. Vague concepts can be approximated using the rough set approach. Uncertainty is the property of elements of a set or a set itself (e.g.,

only examples and/or counterexamples of elements of a considered set are given). Uncertainty of elements of a set can be expressed by the rough membership function.

Both fuzzy and rough set theory represent two different approaches to vagueness. Fuzzy set theory addresses *gradualness* of knowledge, expressed by the fuzzy membership, whereas rough set theory addresses *granularity* of knowledge, expressed by the indiscernibility relation. One can also cope with knowledge gradualness using the rough membership. A nice illustration of this difference was given by Dider Dubois and Henri Prade in their example related to image processing, where fuzzy set theory refers to gradualness of gray level, whereas rough set theory is about the size of pixels.

Consequently, these theories do not compete with each other but are rather complementary. In particular, the rough set approach provides tools for approximate construction of fuzzy membership functions. The rough-fuzzy hybridization approach has proved to be successful in many applications. An interesting discussion of fuzzy and rough set theory in the approach to vagueness can be found in [21.11].

21.9 Discernibility and Boolean Reasoning

The discernibility relations are closely related to indiscernibility relations and belong to the most important relations considered in rough set theory. Tools for discovering and classifying patterns are based on *reasoning schemes* rooted in various paradigms. Such patterns can be extracted from data by means of methods based, e.g., on discernibility and Boolean reasoning.

The ability to discern between perceived objects is important for constructing many entities like reducts, decision rules, or decision algorithms. In the standard approach the discernibility relation $\mathrm{DIS}(B) \subseteq U \times U$ is defined by $x\mathrm{DIS}(B)y$ if and only if $non(x\mathrm{IND}(B)y)$, i.e., $B(x) \cap B(y) = \varnothing$. However, this is, in general, not the case for generalized approximation spaces.

The idea of Boolean reasoning is based on the 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 [21.19–21]. Let us mention that to solve real-life problems it is necessary to deal with very large Boolean functions.

A successful methodology based on the discernibility of objects and Boolean reasoning has been developed for computing many important ingredients for applications. These applications include generation of reducts and their approximations, decision rules, association rules, discretization of real-valued attributes, symbolic value grouping, searching for new features defined by oblique hyperplanes or higher-order surfaces, pattern extraction from data, as well as conflict resolution or negotiation [21.4, 6].

Most of the problems related to the generation of the above-mentioned entities 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 the literature (e.g., with respect to the classification quality of unseen objects). Moreover, they are very efficient from the point of view of the time necessary to compute the solution. Many of these methods are based on discernibility matrices. However, it is possible to compute the necessary information about these matrices without their explicit construction (i.e., by sorting or hashing original data).

It is important to note that the methodology makes it possible to construct heuristics with a very important *approximation property*, which can be formulated as follows: expressions, called *approximate implicants*, generated by heuristics that are *close* to prime implicants define approximate solutions for the problem.

Mining large data sets is one of the biggest challenges in knowledge discovery and data mining (KDD). In many practical applications, there is a need for data mining algorithms running on terminals of a client–server database system where the only access to database (located in the server) is enabled by queries in structured query language (SQL).

Let us consider two illustrative examples of problems for large data sets: (i) searching for short reducts, (ii) searching for best partitions defined by cuts on continuous attributes. In both cases, the traditional implementations of rough sets and Boolean reasoning-based methods are characterized by a high computational cost. The critical factor for the time complexity of algorithms solving the discussed problems is the number of data access operations. Fortunately some efficient modifi-

cations of the original algorithms were proposed by relying on concurrent retrieval of higher-level statistics, which are sufficient for the heuristic search of reducts and partitions [21.4, 6]. The rough set approach was also applied in the development of other scalable big data processing techniques (e.g., [21.22]).

21.10 Rough Sets and Induction

The rough set approach is strongly related to inductive reasoning (e.g., in rough set-based methods for inducing classifiers or clusters [21.6]). The general idea for inducing classifiers is as follows. From a given decision table a set of granules in the form of decision rules is induced together with arguments *for* and *against* each decision rule and decision class. For any new object with known signature one can select rules matching this object. Note that the left-hand sides of decision rules are described by formulae that make it possible to check for new objects if they satisfy them assuming that the signatures of these objects are known. In this way, one can consider two semantics of formulae: on a sample of objects U and on its extension $U^* \supseteq U$. Definitely, one should consider a risk related to such generalization, e.g., in the decision rule induction. Next, a conflict resolution should be applied to resolve conflicts between matched rules by new object voting for different decisions. In the rough set approach, the process of inducing classifiers can be considered as the process of inducing approximations of concepts over extensions of approximation spaces (defined over samples of objects represented by decision systems). The whole procedure can be generalized for the case of approximation of more complex information granules. It is worthwhile mentioning that approaches for inducing approximate reasoning schemes have also been developed.

A typical approach in machine learning is based on inducing classifiers from samples of objects. These classifiers are used for prediction decisions on objects unseen so far, if only the signatures of these objects are available. This approach can be called global, i. e., leading to decision extension from a given sample of objects on the whole universe of objects. This global approach has some drawbacks (see the Epilog in [21.23]). Instead of this one can try to use transduction [21.23], semi-supervised learning, induced local models relative to new objects, or adaptive learning strategies. However, we are still far away from fully understanding the discovery processes behind such generalization strategies [21.24].

21.11 Rough Set–Based Generalizations

The original approach by Pawlak was based on indiscernibility defined by equivalence relations. Any such indiscernibility relation defines a partition of the universe of objects. Over the years many generalizations of this approach were introduced, many of which are based on coverings rather than partitions. In particular, one can consider the similarity (tolerance)-based rough set approach, binary relation based rough sets, neighborhood and covering rough sets, the dominance-based rough set approach, hybridization of rough sets and fuzzy sets, and many others.

One should note that dealing with coverings requires solving several new algorithmic problems such as the selection of family of definable sets or resolving problems with the selection of the relevant definition of the approximation of sets among many possible ones. One should also note that for a given problem (e.g., classification problem) one should discover the relevant covering for the target classification task. In the literature there are numerous papers dedicated to theoretical aspects of the covering rough set approach. However, still much more work should be done on rather hard algorithmic issues, e.g., for the relevant covering discovery.

Another issue to be solved is related to inclusion measures. Parameters of such measures are tuned in inducing of the high quality approximations. Usually, this is done on the basis of the minimum description length principle. In particular, approximation spaces with rough inclusion measures have been investigated. This approach was further extended to the rough mereological approach. More general cases of approximation spaces with rough inclusion have also been discussed in the literature, including approximation spaces in GC.

Finally, it is worthwhile mentioning the approach for ontology approximation used in hierarchical learning of complex vague concepts [21.6].

In this section, we discuss in more detail some issues related to the above-mentioned generalizations. Several generalizations of the classical rough set approach based on approximation spaces defined as pairs of the form (U, R), where R is the equivalence relation (called indiscernibility relation) on the set U, have been reported in the literature. They are related to different views on important components used in the definition of rough sets. In the definition of rough sets different kinds of structural sets that are examples of information granules are used. From mathematical point of view, one may treat them as sets defined over the hierarchy of the powerset of objects. Among them are the following ones:

- Elementary granules (neighborhoods) of objects (e.g., similarity, tolerance, dominance neighborhoods, fuzzy neighborhoods, rough-fuzzy neighborhoods, fuzzy rough neighborhoods, families of neighborhoods).
- Granules defined by accessible information about objects (e.g., only partial information on the signature of objects may be accessible).
- Methods for the definition of higher-order information granules (e.g., defined by the left-hand sides of induced decision rules or clusters of similar information granules).
- Inclusion measures making it possible to define the degrees of inclusion and/or closeness between information granules (e.g., the degrees of inclusion granules defined by accessible information about objects into elementary granules).
- Aggregation methods of inclusion or/and closeness degrees.
- Methods for the definition of approximation operations, including strategies for extension of approximations from samples of objects to larger sets of objects.
- Algebraic structures of approximation spaces.

Let us consider some examples of generalizations of the rough set approach proposed by Pawlak in 1982.

A generalized approximation space [21.25] can be defined by a tuple $\mathcal{AS} = (U, I, \nu)$, where I is the *uncertainty function* defined on U with values in the powerset POW(U) of U ($I(x)$ is the *neighborhood* of x) and ν is the *inclusion function* defined on the Cartesian product POW(U) × POW(U) with values in the interval $[0, 1]$

measuring the degree of inclusion of sets. The lower and upper approximation operations can be defined in $\mathcal{AS}$ by

$$\text{LOW}(\mathcal{AS}, X) = \{x \in U : \nu(I(x), X) = 1\}, \quad (21.11)$$
$$\text{UPP}(\mathcal{AS}, X) = \{x \in U : \nu(I(x), X) > 0\}. \quad (21.12)$$

In the case considered by *Pawlak* [21.2], $I(x)$ is equal to the equivalence class $B(x)$ of the indiscernibility relation IND(B); in the case of the tolerance (or similarity) relation $T \subseteq U \times U$ we take $I(x) = [x]_T = \{y \in U : xTy\}$, i.e., $I(x)$ is equal to the tolerance class of T defined by x.

The standard rough inclusion relation ν_{SRI} is defined for $X, Y \subseteq U$ by

$$\nu_{\text{SRI}}(X, Y) = \begin{cases} \dfrac{\text{card}(X \cap Y)}{\text{card}(X)}, & \text{if } X \neq \emptyset, \\ 1, & \text{otherwise}. \end{cases} \quad (21.13)$$

For applications it is important to have some constructive definitions of I and ν.

One can consider another way to define $I(x)$. Usually together with $\mathcal{AS}$ we consider some set $\mathcal{F}$ of formulae describing sets of objects in the universe U of $\mathcal{AS}$ defined by semantics $\| \cdot \|_{\mathcal{AS}}$, i.e., $\|\alpha\|_{\mathcal{AS}} \subseteq U$ for any $\alpha \in \mathcal{F}$. If $\mathcal{AS} = (U, A)$ then we will also write $\|\alpha\|_U$ instead of $\|\alpha\|_{\mathcal{AS}}$. Now, one can take the set

$$N_{\mathcal{F}}(x) = \{\alpha \in \mathcal{F} : x \in \|\alpha\|_{\mathcal{AS}}\}, \quad (21.14)$$

and $I(x) = \{\|\alpha\|_{\mathcal{AS}} : \alpha \in N_{\mathcal{F}}(x)\}$. Hence, more general uncertainty functions with values in POW(POW(U)) can be defined and in consequence different definitions of approximations are considered. For example, one can consider the following definitions of approximation operations in this approximation space $\mathcal{AS}$

$$\text{LOW}(\mathcal{AS}_\circ, X)$$
$$= \{x \in U : \nu(Y, X) = 1 \text{ for some } Y \in I(x)\}, \quad (21.15)$$
$$\text{UPP}(\mathcal{AS}_\circ, X)$$
$$= \{x \in U : \nu(Y, X) > 0 \text{ for any } Y \in I(x)\}. \quad (21.16)$$

There are also different forms of rough inclusion functions. Let us consider two examples. In the first example of a rough inclusion function, a threshold $t \in (0, 0.5)$ is used to relax the degree of inclusion of sets.

The rough inclusion function v_t is defined by

$$v_t(X, Y)$$
$$= \begin{cases} 1 & \text{if } v_{SRI}(X, Y) \geq 1 - t, \\ \dfrac{v_{SRI}(X, Y) - t}{1 - 2t} & \text{if } t \leq v_{SRI}(X, Y) < 1 - t, \\ 0 & \text{if } v_{SRI}(X, Y) \leq t. \end{cases}$$

(21.17)

One can obtain approximations considered in the variable precision rough set approach (VPRSM) by substituting in (21.12) and (21.13) the rough inclusion function v_t defined by (21.17) instead of v, assuming that Y is a decision class and $I(x) = B(x)$ for any object x, where B is a given set of attributes. Another example of application of the standard inclusion was developed by using probabilistic decision functions. The rough inclusion relation can be also used for function approximation and relation approximation [21.25].

The approach based on inclusion functions has been generalized to the *rough mereological approach* [21.26]. 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 being a generalization of the *Leśniewski* mereology [21.27].

Usually families of approximation spaces labeled by some parameters are considered. By tuning such parameters according to chosen criteria (e.g., minimal description length) one can search for the optimal approximation space for concept approximation.

Our knowledge about the approximated concepts is often partial and uncertain. For example, concept approximation should be constructed from examples and counterexamples of objects for the concepts [21.14]. Hence, concept approximations constructed from a given sample of objects are extended, using inductive reasoning, on objects not yet observed. The rough set approach for dealing with concept approximation under such partial knowledge is now well developed.

Searching strategies for relevant approximation spaces are crucial for real-life applications. They include the discovery of uncertainty functions, inclusion measures, as well as selection of methods for approximations of decision classes and strategies for inductive extension of approximations from samples on larger sets of objects.

Approximations of concepts should be constructed under dynamically changing environments. This leads to a more complex situation where the boundary regions are not crisp sets, which is consistent with the postulate of the higher-order vagueness considered by philosophers [21.10]. Different aspects of vagueness in the rough set framework have been discussed.

It is worthwhile mentioning that a rough set approach to the approximation of compound concepts has been developed. For such concepts, it is hardly possible to expect that they can be approximated with the high quality by the traditional methods [21.23, 28]. The approach is based on hierarchical learning and ontology approximation. Approximation methods of concepts in distributed environments have been developed. The reader may find surveys of algorithmic methods for concept approximation based on rough sets and Boolean reasoning in the literature.

In several papers, the problem of ontology approximation was discussed together with possible applications to approximation of compound concepts or to knowledge transfer. In any ontology [21.29] (vague) concepts and local dependencies between them are specified. Global dependencies can be derived from local dependencies. Such derivations can be used as hints in searching for relevant compound patterns (information granules) in approximation of more compound concepts from the ontology. The ontology approximation problem is one of the fundamental problems related to approximate reasoning in distributed environments. One should construct (in a given language that is different from the language in which the ontology is specified) not only approximations of concepts from ontology, but also vague dependencies specified in the ontology. It is worthwhile mentioning that an ontology approximation should be induced on the basis of incomplete information about concepts and dependencies specified in the ontology. Information granule calculi based on rough sets have been proposed as tools making it possible to solve this problem. Vague dependencies have vague concepts in premises and conclusions.

The approach to approximation of vague dependencies based only on degrees of closeness of concepts from dependencies and their approximations (classifiers) is not satisfactory for approximate reasoning. Hence, more advanced approach should be developed. Approximation of any vague dependency is a method which for any object allows us to compute the arguments *for* and *against* its membership to the dependency conclusion on the basis of analogous arguments relative to the dependency premises. Any argument is a compound information granule (compound pattern). Arguments are fused by local schemes (production rules) discovered from data. Further fusions are

possible through composition of local schemes, called approximate reasoning schemes (AR) [21.30]. To estimate the degree to which (at least) an object belongs to concepts from ontology the arguments *for* and *against* those concepts are collected and next a conflict resolution strategy is applied to them to predict the degree.

Several generalizations of the rough set approach introduced by Pawlak in 1982 are discussed in this handbook in more detail:

- The similarity (tolerance)-based rough set approach (Chap. 25)
- Binary relation based rough sets (Chap. 25)
- Neighborhood and covering rough sets (Chap. 25)
- The dominance-based rough set approach (Chap. 22)
- The probabilistic rough set approach and its probabilistic extension called the variable consistency dominance-based rough set approaches (Chap. 24)
- Parameterized rough sets based on Bayesian confirmation measures (Chap. 22)
- Stochastic rough set approaches (Chap. 22)
- Generalizations of rough set approximation operations (Chap. 25)
- Hybridization of rough sets and fuzzy sets (Chap. 26)
- Rough sets on abstract algebraic structures (e.g., lattices) (Chap. 25).

There are some other well-established or emerging domains not covered in the chapter where some generalizations of rough sets are proposed as the basic tools, often in combination with other existing approaches. Among them are rough sets based on [21.6]:

i) Incomplete information and/or decision systems
ii) Nondeterministic information and/or decision systems
iii) The rough set model on two universes
iv) Dynamic information and/or decision systems
v) Dynamic networks of information and/or decision systems.

Moreover, rough sets play a crucial role in the development of granular computing (GC) [21.31]. The extension to interactive granular computing (IGR) [21.32] requires generalization of basic concepts such as information and decision systems, as well as methods of inducing hierarchical structures of information and decision systems.

Let us note that making progress in understanding interactive computations is one of the key problems in developing high quality intelligent systems working in complex environments [21.33]. The current research projects aim at developing foundations of IGC based on the rough set approach in combination with other soft computing approaches, in particular with fuzzy sets. The approach is called interactive rough granular computing (IRGC). In IRGC computations are based on interactions of complex granules (c-granules, for short). Any c-granule consists of a physical part and a mental part that are linked in a special way [21.32]. IRGC is treated as the basis for (see [21.6] and references in this book):

i) Wistech Technology, in particular for approximate reasoning, called adaptive judgment about properties of interactive computations
ii) Context induction
iii) Reasoning about changes
iv) Process mining (this research was inspired by [21.34])
v) Perception-based computing (PBC)
vi) Risk management in computational systems [21.32].

Interactive computations based on c-granules seem to create a good background, e.g., for modeling computations in Active Media Technology (AMT) and Wisdom Web of Things (W2T). We plan to investigate their role for foundations of natural computing too. Let us also mention that the interactive computations based on c-granules are quite different in nature than Turing computations. Hence, we plan to investigate relationships of interactive computability based on c-granules and Turing computability.

21.12 Rough Sets and Logic

The father of contemporary logic was the German mathematician Gottlob Frege (1848–1925). He thought that mathematics should not be based on the notion of set but on the notions of logic. He created the first axiomatized logical system but it was not understood by the logicians of those days. During the first three

decades of the twentieth century there was a rapid development in logic, bolstered to a great extent by Polish logicians, in particular by Alfred Tarski, Stanisław Leśniewski, Jan Łukasiewicz, and next by Andrzej Mostowski and Helena Rasiowa. The development of computers and their applications stimulated logical research and widened their scope.

When we speak about logic, we generally mean *deductive logic*. It gives us tools designed for deriving true propositions from other true propositions. Deductive reasoning always leads to true conclusions from true premises. The theory of deduction has well-established, generally accepted theoretical foundations. Deductive reasoning is the main tool used in mathematical reasoning.

Rough set theory has contributed to some extent to various kinds of deductive reasoning. Particularly, various kinds of logics based on the rough set approach have been investigated; rough set methodology has contributed essentially to modal logics, many valued logic, intuitionistic logic, and others (see, e.g., references in the book [21.6] and in articles [21.3, 4]). A summary of this research can be found in [21.35, 36] and the interested reader is advised to consult these volumes.

In natural sciences (e.g., in physics) *inductive reasoning* is of primary importance. The characteristic feature of such reasoning is that it does not begin from axioms (expressing general knowledge about the reality) like in deductive logic, but some partial knowledge (examples) about the universe of interest are the starting point of this type of reasoning, which are generalized next and they constitute the knowledge about a wider reality than the initial one. In contrast to deductive reasoning, inductive reasoning does not lead to true conclusions but only to probable (possible) ones. Also, in contrast to the logic of deduction, the logic of induction does not have uniform, generally accepted, theoretical foundations as yet, although many important and interesting results have been obtained, e.g., concerning statistical and computational learning and others.

Verification of the validity of hypotheses in the logic of induction is based on experiment rather than the formal reasoning of the logic of deduction. Physics is the best illustration of this fact. The research on modern inductive logic has a several centuries' long history. It is worthwhile mentioning here the outstanding English philosophers Francis Bacon (1561–1626) and *John Stuart Mill* (1806–1873) [21.37].

The creation of computers and their innovative applications essentially contributed to the rapid growth of interest in inductive reasoning. This domain is developing very dynamically thanks to computer science. Machine learning, knowledge discovery, reasoning from data, expert systems, and others are examples of new directions in inductive reasoning. Rough set theory is very well suited as a theoretical basis for inductive reasoning. Basic concepts of this theory fit very well to represent and analyze knowledge acquired from examples, which can be next used as a starting point for generalization. Besides, in fact, rough set theory has been successfully applied in many domains to find patterns in data (data mining) and acquire knowledge from examples (learning from examples). Thus, rough set theory seems to be another candidate as a mathematical foundation of inductive reasoning.

The most interesting from a computer science point of view is *common sense* reasoning. We use this kind of reasoning in our everyday lives, and we face examples of such kind of reasoning in newspapers, radio, TV, etc., in political, economics, etc., and in debates and discussions.

The starting point for such reasoning is the knowledge possessed by a specific group of people (*common knowledge*) concerning some subject and intuitive methods of deriving conclusions from it. Here we do not have the possibility to resolve the dispute by means of methods given by deductive logic (reasoning) or by inductive logic (experiment). So the best known methods for solving the dilemma are voting, negotiations, or even war. See, e.g., Gulliver's Travels [21.38], where the hatred between Tramecksan (High-Heels) and Slamecksan (Low-Heels) or disputes between Big-Endians and Small-Endians could not be resolved without a war. These methods do not reveal the truth or falsity of the thesis under consideration at all. Of course, such methods are not acceptable in mathematics or physics. Nobody is going to solve the truth of Fermat's theorem or Newton's laws by voting, negotiations, or declare a war.

Reasoning of this kind is the least studied from the theoretical point of view and its structure is not sufficiently understood, in spite of many interesting theoretical research in this domain [21.39]. The meaning of commonsense reasoning, considering its scope and significance for some domains, is fundamental, and rough set theory can also play an important role in it, but more fundamental research must be done to this end. In particular, the rough truth introduced and studied in [21.40] seems to be important for investigating commonsense reasoning in the rough set framework.

Let us consider a simple example. In the decision system considered we assume $U = $ Birds is a set of birds that are described by some condition attributes from a set A. The decision attribute is a binary attribute Flies with possible values *yes* if the given bird flies and *no*, otherwise. Then, we define the set of abnormal birds by $\text{Ab}_A(\text{Birds}) = A_*(\{x \in \text{Birds}: \text{Flies}(x) = no\})$. Hence, we have, $\text{Ab}_A(\text{Birds}) = \text{Birds} - A^*(\{x \in \text{Birds}: \text{Flies}(x) = yes\})$ and $\text{Birds} - Ab_A(\text{Birds}) = A^*(\{x \in \text{Birds}: \text{Flies}(x) = yes\})$. This means that for normal birds it is consistent, with knowledge represented by A, to assume that they can fly, i.e., it is possible that they can fly. One can optimize $Ab_A(\text{Birds})$ using A to obtain minimal boundary region in the approximation of $\{x \in \text{Birds}: \text{Flies}(x) = no\}$.

It is worthwhile mentioning that in [21.41] an approach was presented that combines the rough sets with nonmonotonic reasoning. Some basic concepts are distinguished, which can be approximated on the basis of sensor measurements and more complex concepts that are approximated using so-called transducers defined by first-order theories constructed over approximated concepts. Another approach to commonsense reasoning was developed in a number of papers. The approach is based on an ontological framework for approximation. In this approach, approximations are constructed for concepts and dependencies between the concepts represented in a given ontology, expressed, e.g., in natural language. Still another approach combining rough sets with logic programming has been developed. Let us also note that *Pawlak* proposed a new approach to conflict analysis [21.42]. The approach was next extended in the rough set framework.

To recapitulate, let us consider the following characteristics of the three above-mentioned kinds of reasoning:

a) *Deductive*
   i) *Reasoning methods*: axioms and rules of inference
   ii) *Applications*: mathematics
   iii) *Theoretical foundations*: complete theory
   iv) *Conclusions*: true conclusions from true premisses
   v) *Hypotheses verification*: formal proof
b) *Inductive*
   i) *Reasoning methods*: generalization from examples
   ii) *Applications*: natural sciences (physics)
   iii) *Theoretical foundations*: lack of generally accepted theory

iv) *Conclusions*: not true but probable (possible)
   v) *Hypotheses verification*: empirical experiment
c) *Common sense*
   i) *Reasoning methods*: reasoning method based on common sense knowledge with intuitive rules of inference expressed in natural language
   ii) *Applications*: everyday life, humanities
   iii) *Theoretical foundations*: lack of generally accepted theory
   iv) *Conclusions*: obtained by mixture of deductive and inductive reasoning based on concepts expressed in natural language, e.g., with application of different inductive strategies for conflict resolution (such as voting, negotiations, cooperation, war) based on human behavioral patterns
   v) *Hypotheses verification*: human behavior.

There are numerous issues related to approximate reasoning under uncertainty. These issues are discussed in books on granular computing, rough mereology, and the computational complexity of algorithmic problems related to these issues. For more details, the reader is referred to the following books [21.26, 31, 43, 44].

Finally, we would like to stress that still much more work should be done to develop approximate reasoning about complex vague concepts to make progress in the development of intelligent systems. According to *Leslie Valiant* [21.45] (who is the 2011 winner of the ACM Turing Award, for his fundamental contributions to the development of computational learning theory and to the broader theory of computer science):

A fundamental question for artificial intelligence is to characterize the computational building blocks that are necessary for cognition. A specific challenge is to build on the success of machine learning so as to cover broader issues in intelligence ... This requires, in particular a reconciliation between two contradictory characteristics – the apparent logical nature of reasoning and the statistical nature of learning.

It is worthwhile presenting two more views. The first one by *Lotfi A. Zadeh*, the founder of fuzzy sets and the computing with words (CW) paradigm [21.46]:

Manipulation of perceptions plays a key role in human recognition, decision and execution processes. As a methodology, computing with words provides a foundation for a computational theory of per-

ceptions – a theory which may have an important bearing on how humans make and machines might make – perception-based rational decisions in an environment of imprecision, uncertainty and partial truth. . . . computing with words, or CW for short, is a methodology in which the objects of computation are words and propositions drawn from a natural language.

The other view is that of *Judea Pearl* [21.47] (the 2011 winner of the ACM Turing Award, the highest distinction in computer science, for fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning):

Traditional statistics is strong in devising ways of describing data and inferring distributional parameters from sample. Causal inference requires two additional ingredients: a science-friendly language for articulating causal knowledge, and a mathematical machinery for processing that knowledge, combining it with data and drawing new causal conclusions about a phenomenon.

The question arises about the logic relevant for the above-mentioned tasks. First, let us observe that the satisfiability relations in the IRGC framework can be treated as tools for constructing new granules. In fact, for a given satisfiability relation one can define the semantics of formulae related to this relation, i. e., which are the candidates for the new relevant granules. We would like to emphasize one a very important feature. The relevant satisfiability relation for the considered problems is not given but it should be induced (discovered) from partial information given by information or decision systems. For real-life problems it is often necessary to discover a hierarchy of satisfiability relations before we obtain the relevant target one. Granules constructed on different levels of this hierarchy finally lead to relevant ones for approximation of complex vague concepts related to complex granules expressed using natural language.

The reasoning making it possible to derive relevant c-granules for solutions of the target tasks is called *adaptive judgment*. *Intuitive judgment* and *ra-*

tional judgment are distinguished as different kinds of judgment [21.48]. Deduction and induction as well as abduction or analogy-based reasoning are involved in adaptive judgment. Among the tasks for adaptive judgment are the following ones, which support reasoning under uncertainty toward: searching for relevant approximation spaces, discovery of new features, selection of relevant features, rule induction, discovery of inclusion measures, strategies for conflict resolution, adaptation of measures based on the minimum description length principle, reasoning about changes, perception (action and sensory) attributes' selection by agent control, adaptation of quality measures over computations relative to agents, adaptation of object structures, discovery of relevant contexts, strategies for knowledge representation and interaction with knowledge bases, ontology acquisition and approximation, learning in dialog of inclusion measures between granules from different languages (e.g., the formal language of the system and the user's natural language), strategies for adaptation of existing models, strategies for development and evolution of communication language among agents in distributed environments, strategies for risk management in distributed computational systems. Definitely, in the language used by agents for dealing with adaptive judgment (i. e., intuitive and rational) some deductive systems known from logic may be applied for reasoning about knowledge relative to closed worlds. This may happen, e.g., if the agent languages are based on classical mathematical logic. However, if we move to interactions in open worlds, then new specific rules or patterns relative to a given agent or group of agents in such worlds should be discovered. The process of inducing such rules or patterns is influenced by uncertainty because they are induced by agents under uncertain and/or imperfect knowledge about the environment.

The concepts discussed, such as interactive computation and adaptive judgment, are among the basic concepts in Wisdom Technology (WisTech) [21.49, 50]. Let us mention here the WisTech meta-equation

$$\begin{aligned} \text{WISDOM} = \ &\text{INTERACTIONS} \\ &+ \text{ADAPTIVE JUDGMENT} \\ &+ \text{KNOWLEDGE} \,. \end{aligned} \tag{21.18}$$

21.13 Conclusions

In the chapter, we have discussed some basic issues and methods related to rough sets together with some generalizations, including those related to relationships of rough sets with inductive reasoning. We have also listed some current research directions based on interactive rough granular computing. For more details, the reader is referred to the literature cited at the beginning of this chapter (see also [21.9]).

References

21.1   Z. Pawlak: *Rough Sets: Theoretical Aspects of Reasoning about Data*, Theory and Decision Library D, Vol. 9 (Kluwer, Dordrecht 1991)

21.2   Z. Pawlak: Rough sets, Int. J. Comp. Inform. Sci. **11**, 341–356 (1982)

21.3   Z. Pawlak, A. Skowron: Rudiments of rough sets, Inform. Sci. **177**(1), 3–27 (2007)

21.4   Z. Pawlak, A. Skowron: Rough Sets: Some Extensions, Inform. Sci. **177**(1), 28–40 (2007)

21.5   Z. Pawlak, A. Skowron: Rough sets and Boolean reasoning, Inform. Sci. **177**(1), 41–73 (2007)

21.6   A. Skowron, Z. Suraj (Eds.): *Rough Sets and Intelligent Systems. Professor Zdzislaw Pawlak in Memoriam*, Intelligent Systems Reference Library, Vol. 42/43 (Springer, Berlin, Heidelberg 2013)

21.7   I. Chikalov, V. Lozin, I. Lozina, M. Moshkov, H.S. Nguyen, A. Skowron, B. Zielosko: *Three Approaches to Data Analysis. Test Theory, Rough Sets and Logical Analysis of Data*, Intelligent Systems Reference Library, Vol. 41 (Springer, Berlin, Heidelberg 2012)

21.8   Ch. Cornelis (Ed.): International Rough Sets Society, online available from: http://www.roughsets.org

21.9   Z. Suraj: Rough Set Database System, online available from: http://www.rsds.univ.rzeszow.pl

21.10   R. Keefe: *Theories of Vagueness*, Cambridge Studies in Philosophy (Cambridge Univ. Press, Cambridge 2000)

21.11   S. Read: *Thinking about Logic: An Introduction to the Philosophy of Logic* (Oxford Univ. Press, Oxford 1994)

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

21.13   G.W. Leibniz: Discourse on Metaphysics. In: *Philosophical Essays (1686)*, ed. by R. Ariew, D. Garber (Hackett, Indianapolis 1989) pp. 35–68

21.14   T. Hastie, R. Tibshirani, J.H. Friedman: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, Berlin, Heidelberg 2001)

21.15   J. Rissanen: Modeling by shortes data description, Automatica **14**, 465–471 (1978)

21.16   J. Łukasiewicz: Die logischen Grundlagen der Wahrscheinlichkeitsrechnung. In: Jan Łukasiewicz – Selected Works, ed. by L. Borkowski (North Holland/Polish Scientific Publishers, Amsterdam, Warsaw 1970) pp. 16–63

21.17   J. Barwise, J. Seligman: *Information Flow: The Logic of Distributed Systems* (Cambridge Univ. Press, Cambridge 1997)

21.18   L.A. Zadeh: Fuzzy sets, Inform. Control **8**, 338–353 (1965)

21.19   G. Boole: *The Mathematical Analysis of Logic* (G. Bell, London 1847), Reprinted by Philosophical Library, New York 1948

21.20   G. Boole: *An Investigation of the Laws of Thought* (Walton, London 1954)

21.21   F.M. Brown: *Boolean Reasoning* (Kluwer, Dordrecht 1990)

21.22   D. Slezak: Infobright, online available from: http://www.infobright.com/

21.23   V. Vapnik: *Statistical Learning Theory* (Wiley, New York 1998)

21.24   R.S. Michalski: A theory and methodology of inductive learning, Artif. Intell. **20**, 111–161 (1983)

21.25   J. Stepaniuk: *Rough-Granular Computing in Knowledge Discovery and Data Mining* (Springer, Berlin, Heidelberg 2008)

21.26   L. Polkowski: *Approximate Reasoning by Parts. An Introduction to Rough Mereology*, Intelligent Systems Reference Library, Vol. 20 (Springer, Berlin, Heidelberg 2011)

21.27   S. Leśniewski: Grungzüge eines neuen Systems der Grundlagen der Mathematik, Fundam. Math. **14**, 1–81 (1929)

21.28   L. Breiman: Statistical modeling: The two cultures, Stat. Sci. **16**(3), 199–231 (2001)

21.29   S. Staab, R. Studer: *Handbook on Ontologies*, International Handbooks on Information Systems (Springer, Berlin 2004)

21.30   S.K. Pal, L. Polkowski, A. Skowron: *Rough-Neural Computing: Techniques for Computing with Words*, Cognitive Technologies (Springer, Berlin, Heidelberg 2004)

21.31   W. Pedrycz, S. Skowron, V. Kreinovich (Eds.): *Handbook of Granular Computing* (Wiley, Hoboken 2008)

21.32   A. Jankowski: *Practical Issues of Complex Systems Engineering: Wisdom Technology Approach* (Springer, Berlin, Heidelberg 2015), in preparation

21.33   D. Goldin, S. Smolka, P. Wegner: *Interactive Computation: The New Paradigm* (Springer, Berlin, Heidelberg 2006)

21.34 Z. Pawlak: Concurrent versus sequential – the rough sets perspective, Bulletin EATCS **48**, 178–190 (1992)

21.35 L. Polkowski: *Rough Sets: Mathematical Foundations*, Advances in Soft Computing (Physica, Berlin, Heidelberg 2002)

21.36 M. Chakraborty, P. Pagliani: *A Geometry of Approximation: Rough Set Theory – Logic, Algebra and Topology of Conceptual Patterns* (Springer, Berlin, Heidelberg 2008)

21.37 D.M. Gabbay, S. Hartmann, J. Woods (Eds.): *Inductive Logic*, Handbook of the History of Logic, Vol. 10 (Elsevier, Amsterdam 2011)

21.38 J. Swift: *Gulliver's Travels into Several Remote Nations of the World* (anonymous publisher, London 1726)

21.39 D.M. Gabbay, C.J. Hogger, J.A. Robinson: *Nonmonotonic Reasoning and Uncertain Reasoning*, Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 3 (Clarendon, Oxford 1994)

21.40 Z. Pawlak: Rough logic, Bull. Pol. Ac.: Tech. **35**(5/6), 253–258 (1987)

21.41 P. Doherty, W. Łukaszewicz, A. Skowron, A. Szałas: *Knowledge Engineering: A Rough Set Approach*, Studies in Fizziness and Soft Computing, Vol. 202 (Springer, Berlin, Heidelberg 2006)

21.42 Z. Pawlak: An inquiry into anatomy of conflicts, Inform. Sci. **109**, 65–78 (1998)

21.43 M.J. Moshkov, M. Piliszczuk, B. Zielosko: *Partial Covers, Reducts and Decision Rules in Rough Sets – Theory and Applications*, Studies in Computational Intelligence, Vol. 145 (Springer, Berlin, Heidelberg 2008)

21.44 P. Delimata, M.J. Moshkov, A. Skowron, Z. Suraj: *Inhibitory Rules in Data Analysis: A Rough Set Approach*, Studies in Computational Intelligence, Vol. 163 (Springer, Berlin, Heidelberg 2009)

21.45 Web page of Professor Leslie Valiant, online available from: http://people.seas.harvard.edu/~valiant/researchinterests.htm

21.46 L.A. Zadeh: From computing with numbers to computing with words – From manipulation of measurements to manipulation of perceptions, IEEE Trans. Circuits Syst. **45**, 105–119 (1999)

21.47 J. Pearl: Causal inference in statistics: An overview, Stat. Surv. **3**, 96–146 (2009)

21.48 D. Kahneman: Maps of Bounded Rationality: Psychology for behavioral economics, Am. Econ. Rev. **93**, 1449–1475 (2002)

21.49 A. Jankowski, A. Skowron: A wistech paradigm for intelligent systems, Lect. Notes Comput. Sci. **4374**, 94–132 (2007)

21.50 A. Jankowski, A. Skowron: Logic for artificial intelligence: The Rasiowa–Pawlak school perspective. In: *Andrzej Mostowski and Foundational Studies*, ed. by A. Ehrenfeucht, V. Marek, M. Srebrny (IOS, Amsterdam 2008) pp. 106–143