Chapter 5

Rough Sets

5.1 Introduction

The methodology we propose and develop in this book is founded on the notion of rough sets. In many AI applications one faces the problem of representing and processing incomplete, imprecise, and approximate data. Many of these applications require the use of approximate reasoning techniques. Before we introduce rough sets formally, let us begin with an intuitive example where representation of approximate data and reasoning with it is an essential component in the modeling process.

Example 5.1.1 Consider a UAV equipped with a sensor platform which includes a digital camera. Suppose that the UAV task is to recognize various situations on roads. It is assumed that the camera has a particular resolution. It follows that the precise shape of the road cannot be recognized if essential features of the road shape require a higher resolution then that provided by the camera. Figure 5.1 depicts a view from the UAV's camera, where a fragment of a road is shown together with three cars c1, c2, and c3.

Observe that due to the camera resolution there are collections of points that should be interpreted as being indiscernible with each other. The collections of indiscernible points are called *elementary sets*, using rough set terminology. In Figure 5.1, elementary sets are illustrated by dashed squares and correspond to pixels. Any point in a pixel is not discernable from any other point in the pixel from the perspective of the UAV. Elementary sets are then used to approximate objects that cannot be precisely represented by means of (unions of) elementary sets. For instance, in Figure 5.1, it can be observed that for some elementary sets one part falls within and the other outside the actual road boundaries (represented by curved lines).

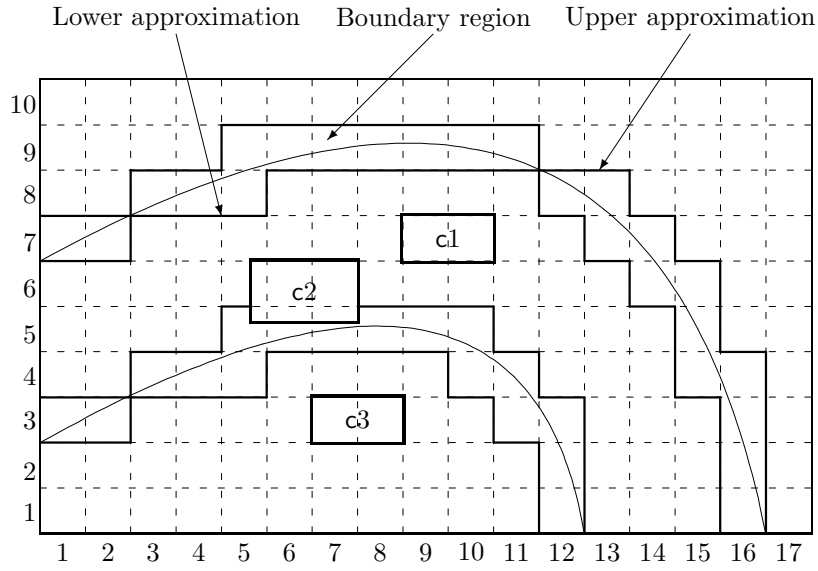

Figure 5.1: Sensing a road considered in Example 5.1.1.

Instead of a precise characterization of the road and cars, using rough set techniques, one can obtain approximate characterizations as depicted in Figure 5.2. Observe that the road sequence is characterized only in terms of a lower and upper approximation of the actual road. A boundary region, containing points that are unknown to be inside or outside of the road's boundaries, is characterized by a collection of elementary sets marked with dots inside. Cars c1 and c3 are represented precisely, while car c2 is represented by its lower approximation (the thick box denoted by c2) and by its upper approximation (the lower approximation together with the region containing elementary sets marked by hollow dots inside). The region of elementary sets marked by hollow dots inside represents the boundary region of the car.

The approximations of the concepts are based on available information about points expressed by means of pixels. The lower approximation of a concept represents points that are known to be part of the concept, the boundary region represents points that might or might not be part of the concept, and the complement of the upper approximation represents points that are known not to be part of the concept. Consequently, car c1 is characterized as being completely on the road (inside the roads boundaries); it is unknown whether car c2 is completely on the road and car c3 is known to be outside, or off the road. ■

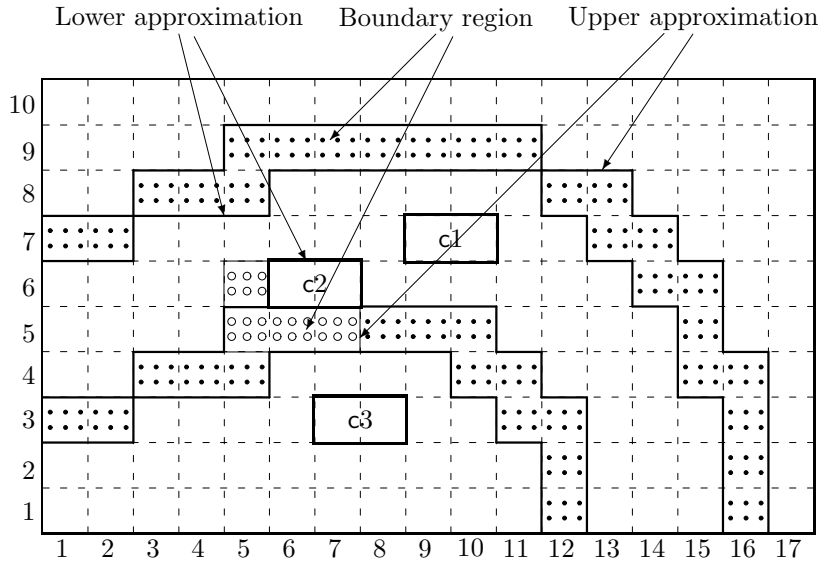

Figure 5.2: The approximate view of the road considered in Example 5.1.1.

As illustrated in Example 5.1.1, the rough set philosophy is founded on the assumption that we associate some information (data, knowledge) with every object of the universe of discourse. This information is often formulated in terms of attributes about objects. Objects characterized by the same information are interpreted as indiscernible (similar) in view of the available information about them. An indiscernibility relation, generated in this manner from the attribute/value pairs associated with objects, provides the mathematical basis of rough set theory.

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 in a universe is referred to as a *crisp set*; otherwise the set is referred to as being a *rough set* . In the latter case, two separate unions of elementary sets can be used to approximate the imprecise set, as we have seen in the example above. Since a relation is a set of tuples, *rough relation* is defined to be a rough set of tuples.

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

The assumption that objects can be *observed* only through the information available about them leads to the view that knowledge about objects has granular structure. Due to this granularity, some objects of interest cannot always be discerned given the information available, therefore the objects appear as the same (or similar). As a consequence, vague or imprecise concepts, in contrast to precise concepts, cannot be characterized solely in terms of information about their elements since elements are not always discernable from each other. In the proposed approach, we assume that any vague or imprecise concept is replaced by a pair of precise concepts called the lower and the upper approximation of the vague or imprecise concept. The lower approximation consists of all objects which with certainty belong to the concept and the upper approximation consists of all objects which have a possibility of belonging to the concept.

The difference between the upper and the lower approximation constitutes the boundary region of a vague or imprecise concept. Additional information about attribute values of objects classified as being in the boundary region of a concept may result in such objects being re-classified as members of the lower approximation or as not being included in the concept. Upper and lower approximations are two of the basic operations in rough set theory.

5.2 Approximations

5.2.1 The Basic Concepts

We now introduce the concept of approximations more formally.

Let U be a set of objects. Any partition $\mathcal{E} = \{E_i : i \in I\}$ of U can be considered as a family of *elementary sets*. Of course, the choice of the family depends on a particular application.

Definition 5.2.1 Let U be a set of objects, $\mathcal{E} = \{E_i : i \in I\}$ be a family of elementary sets and $X \subseteq U$. The *lower approximation* and *upper approximation* of X wrt $\mathcal{E}$, denoted by $X_{\mathcal{E}^+}$ and $X_{\mathcal{E}^\oplus}$ respectively, are defined by $X_{\mathcal{E}^+} = \bigcup_{E_i \subseteq X} E_i$ and $X_{\mathcal{E}^\oplus} = \bigcup_{E_i \cap X \neq \emptyset} E_i$. ∎

Example 5.2.2 Let U be the set of non-negative reals. Elementary sets can be defined as intervals: $\mathcal{E} = \{[i, i+1) : i \in \omega\}$. Let $X = [2.4, 7.2]$. Then $X_{\mathcal{E}^+} = [3.0, 7.0)$ and $X_{\mathcal{E}^\oplus} = [2.0, 8.0]$. ∎

5.2.2 Representing Approximations in Logic

Approximations of sets and relations can be expressed in logic.

In order to construct a language for dealing with rough concepts, we introduce the following relation symbols for any rough relation R (see Figure 5.3):

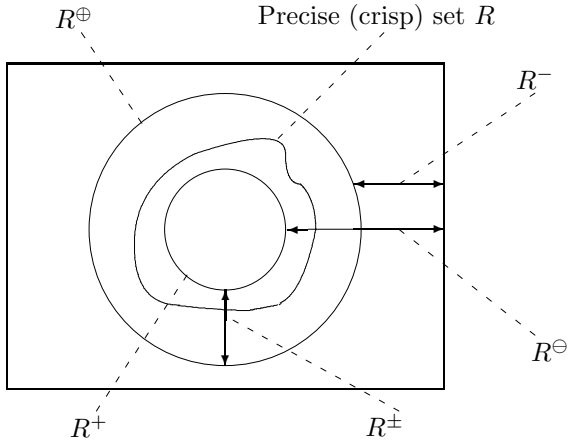

Figure 5.3: Representation of a rough set in logic.

- R^{+} – represents the positive facts known about the relation. R^{+} corresponds to the lower approximation of R. R^{+} is called the *positive region* (*part*) of R.
- R^{-} – represents the negative facts known about the relation. R^{-} corresponds to the complement of the upper approximation of R. R^{-} is called the *negative region* (*part*) of R.
- $R^{\pm}$ – represents the unknown facts about the relation. $R^{\pm}$ corresponds to the set difference between the upper and lower approximations to R. $R^{\pm}$ is called the *boundary region* (*part*) of R.
- $R^{\oplus}$ – represents the positive facts known about the relation together with the unknown facts. $R^{\oplus}$ corresponds to the upper approximation to R. $R^{\oplus}$ is called the *positive-boundary region* (*part*) of R.
- $R^{\ominus}$ – represents the negative facts known about the relation together with the unknown facts. $R^{\ominus}$ corresponds to the lower approximation of the complement of R. $R^{\ominus}$ is called the *negative-boundary region* (*part*) of R.

¿From the logical point of view, elementary sets can be represented by means of logical formulas or primitive relations, assuming their extensions form a

partition of the universe[1]. Assume we are given elementary sets defined by formulas $\{\alpha_1(\bar{x}_1), \ldots, \alpha_n(\bar{x}_n)\}$. Any relation can now be approximated as follows:

$$R^{+}(\bar{x}) \stackrel{\text{def}}{\equiv} \bigvee_{1 \leq i \leq n} \{\alpha_i : \forall \bar{x} \forall \bar{x}_i.(\alpha_i(\bar{x}_i) \rightarrow R(\bar{x}))\}$$

$$R^{\oplus}(\bar{x}) \stackrel{\text{def}}{\equiv} \bigvee_{1 \leq j \leq n} \{\alpha_j : \exists \bar{x} \exists \bar{x}_j.(R(\bar{x}) \wedge \alpha_j(\bar{x}_j))\}.$$

5.3 Information Systems and Indiscernibility

One of the basic fundaments of rough set theory is the indiscernibility relation which is generated using information about particular objects of interest. Information about objects is represented in the form of a set of attributes and their associated values for each object. The indiscernibility relation is intended to express the fact that, due to lack of knowledge, we are unable to discern some objects from others simply by employing the available information about those objects. In general, this means that instead of dealing with each individual object we often have to consider clusters of indiscernible objects as fundamental concepts of our theories.

Let us now present this intuitive picture about rough set theory more formally.

Definition 5.3.1 An *information system* is any pair $\mathcal{A} = \langle U, A \rangle$, where U is a non-empty finite set of *objects*, called the *universe*, and A is a non-empty finite set of functions, called *attributes*, such that $a : U \rightarrow V_a$ for every $a \in A$. The set V_a is called the *value set* of a. By $Inf_B(x) = \{\langle a, a(x) \rangle : a \in B\}$, we denote the *information signature* of x wrt B, where $B \subseteq A$ and $x \in U$. ∎

Information systems are often represented in a form of tables with the first column containing objects and the remaining columns, separated by vertical double lines, containing values of attributes. Such tables are called *information tables*.

Example 5.3.2 Consider an information system $\mathcal{A} = \langle U, A \rangle$, where

- $U = \{\mathsf{car1}, \mathsf{car2}, \mathsf{car3}\}$
- $A = \{Size, Color\}$; $Size(\mathsf{car1}) = Size(\mathsf{car2}) = \mathsf{large}$, $Size(\mathsf{car3}) = \mathsf{small}$; $Color(\mathsf{car1}) = Color(\mathsf{car3}) = \mathsf{red}$, $Color(\mathsf{car2}) = \mathsf{blue}$.

[1] In practice it is sufficient to require that the extensions form a covering of the universe - see Chapter 14.

Object	*Size*	*Color*
car1	large	red
car2	large	blue
car3	small	red

Table 5.1: Information table considered in Example 5.3.2.

The information table corresponding to $\mathcal{A}$ is represented in Table 5.1. ■

Note that in this definition, attributes are treated as functions on objects, where $a(x)$ denotes the value the object x has for the attribute a.

Any subset B of A determines a binary relation $IND_{\mathcal{A}}(B) \subseteq U \times U$, called an indiscernibility relation, defined as follows.

Definition 5.3.3 Let $\mathcal{A} = \langle U, A\rangle$ be an information system and let $B \subseteq A$. By the *indiscernibility relation determined by* B, denoted by $IND_{\mathcal{A}}(B)$, we understand the equivalence relation

$$IND_{\mathcal{A}}(B) = \{\langle x, x'\rangle \in U \times U : \forall a \in B.[a(x) = a(x')]\}.$$

If $\langle x, y\rangle \in IND_{\mathcal{A}}(B)$ we say that x and y are B-*indiscernible*. Equivalence classes of the relation $IND_{\mathcal{A}}(B)$, denoted by $[x]_B$, are referred to as B-*elementary sets*. The unions of B-elementary sets are called B-*definable sets*. ■

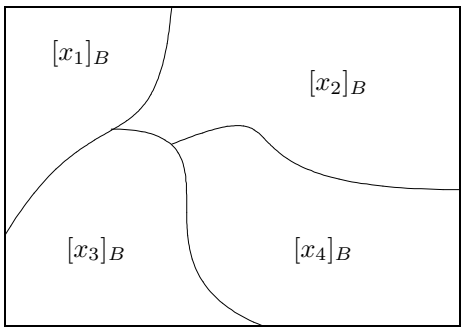

Figure 5.4: A partition $IND_{\mathcal{A}}(B)$.

Of course, $IND_{\mathcal{A}}(B)$ is an equivalence relation. By U/B we denote the set of equivalence classes of relation $IND_{\mathcal{A}}(B)$ and by $[x]_B$ we denote the equivalence class containing x. For example, in Figure 5.4, the partition of

U defined by an indiscernibility relation $IND_{\mathcal{A}}(B)$ contains four equivalence classes, $[x_1]_B, [x_2]_B, [x_3]_B$ and $[x_4]_B$. An example of a B-definable set would be $[x_1]_B \cup [x_4]_B$, where $[x_1]_B$ and $[x_4]_B$ are B-elementary sets.

In Example 5.1.1 the indiscernibility relation is defined by a partition corresponding to pixels represented in Figures 5.1 and 5.2 by squares with dashed borders. Each square represents an elementary set. In the rough set approach the elementary sets are the basic building blocks (concepts) of our knowledge about reality.

The ability to discern between perceived objects is also important for constructing many entities like reducts, decision rules, or decision algorithms which will be considered in later chapters (see Chapters 15, 13).

In the classical rough set approach the *discernibility relation*, $DIS_{\mathcal{A}}(B)$, is defined as follows.

Definition 5.3.4 Let $\mathcal{A} = \langle U, A \rangle$ be an information system and $B \subseteq A$. The *discernibility relation* $DIS_{\mathcal{A}}(B) \subseteq U \times U$ is defined by $xDIS(B)y$ if and only if $\neg(xIND(B)y)$. ∎

We now consider how to define sets of objects using formulas constructed from attribute/value pairs.

Definition 5.3.5 An *elementary descriptor* (*descriptor*, for short) is any expression of the form $(a = v)$, where $a \in A$ and $v \in V_a$. A *generalized descriptor* is any formula of the form $\bigvee_{i=1}^{n}(a = v_i)$, where $a \in A$ and each $v_i \in V_a$. A *Boolean descriptor* is any Boolean combination of elementary descriptors. A *template* is a conjunction of elementary descriptors. ∎

Strictly speaking, a descriptor $(a = v)$ should be written as a relational expression of the form $\lambda x.a(x) = v$, but we shall stick to the simplified form as in the above definition.

Let φ be a Boolean descriptor. The meaning of φ in $\mathcal{A}$, i.e., the set of all objects satisfying φ in $\mathcal{A}$, denoted $\|\varphi\|_{\mathcal{A}}$, is defined inductively as follows:

1. if φ is of the form $(a = v)$ then $\|\varphi\|_{\mathcal{A}} = \{x \in U : a(x) = v\}$;
2. $\|\varphi \wedge \varphi'\|_{\mathcal{A}} = \|\varphi\|_{\mathcal{A}} \cap \|\varphi'\|_{\mathcal{A}}$
   $\|\varphi \vee \varphi'\|_{\mathcal{A}} = \|\varphi\|_{\mathcal{A}} \cup \|\varphi'\|_{\mathcal{A}}$
   $\|\neg\varphi\|_{\mathcal{A}} = U - \|\varphi\|_{\mathcal{A}}$.

Definition 5.3.6 We say that a set of objects $X \subseteq U$ is definable in $\mathcal{A}$ by some formula φ iff $X = \|\varphi\|_{\mathcal{A}}$.

Any $X \subseteq U$ is definable in $\mathcal{A}$ is referred to as a *crisp* (*precise, exact*) set; otherwise the set is referred to as a *rough* (*vague, imprecise, inexact*) set. ∎

5.4 Approximations in Information Systems

Let us now define approximations of sets in the context of information systems.

Definition 5.4.1 Let $\mathcal{A} = \langle U, A \rangle$ be an information system, $B \subseteq A$ and $X \subseteq U$. The *B-lower approximation* and *B-upper approximation* of X, denoted by X_{B^+} and $X_{B^\oplus}$ respectively, are defined by $X_{B^+} = \{x : [x]_B \subseteq X\}$ and $X_{B^\oplus} = \{x : [x]_B \cap X \neq \emptyset\}$. ∎

Definition 5.4.2 The set consisting of objects in the *B-lower approximation* X_{B^+} is also called the *B-positive region* of X. The set $X_{B^-} = U - X_{B^\oplus}$ is called the *B-negative region* of X. The set $X_{B^\pm} = X_{B^\oplus} - X_{B^+}$ is called the *B-boundary region* of X. ∎

Observe that the positive region of X consists of objects that can be classified with certainty as belonging to X using attributes from B. The negative region of X consists of those objects which can be classified with certainty as not belonging to X using attributes from B. The B-boundary region of X consists of those objects that cannot be classified neither as belonging to X nor as not belonging to X, using attributes from B.

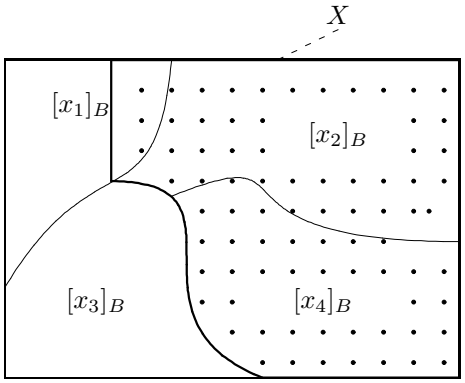

Figure 5.5: A partition $IND_{\mathcal{A}}(B)$ and an imprecise set X.

For example, in Figure 5.5, The B-lower approximation of the set X, is $X_{B^+} = [x_2]_B \cup [x_4]_B$. The B-upper approximation, is $X_{B^\oplus} = [x_1]_B \cup [x_2]_B \cup [x_4]_B \equiv [x_1]_B \cup X_{B^+}$. The B-boundary region, is $X_{B^\pm} = [x_1]_B$. The B-negative region of X, is $X_{B^-} = [x_3]_B \equiv U - X_{B^\oplus}$.

The size of the boundary region of a set can be used as a measure of the quality of that set's approximation. One such measure is defined as follows.

Definition 5.4.3 The *accuracy of approximation* is defined in terms of the following coefficient:

$$\alpha_B(X) \stackrel{\text{def}}{=} \frac{|X_{B^+}|}{|X_{B^\oplus}|},$$

where $|X|$ denotes the cardinality of $X \neq \emptyset$. ∎

It is clear that $0 \leq \alpha_B(X) \leq 1$. If $\alpha_B(X) = 1$ then X is crisp with respect to B (X is precise with respect to B); otherwise, if $\alpha_B(X) < 1$ then X is rough with respect to B (X is vague or imprecise with respect to B).

5.5 Rough Sets and Membership Functions

Rough membership functions can also be introduced in the context of rough set theory.

Definition 5.5.1 Given a set of attributes B and a set X, a *rough membership function* μ_X^B is defined as

$$\mu_X^B(x) \stackrel{\text{def}}{=} \frac{|X \cap [x]_B|}{|[x]_B|}.$$ ∎

It is clear that $0 \leq \mu_X^B(x) \leq 1$. The number $\mu_X^B(x)$ provides a ratio between how much of $[x]_B$ is in X and the cardinality of $[x]_B$. This is depicted in Figure 5.6.

The membership function $\mu_X(x)$ is similar to a conditional probability and its value can be interpreted as the degree of certainty to which x belongs to X. In other words, given X, what is the degree of certainty to which x belongs to X?

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

$$\begin{aligned} X_{B^+} &= \{x \in U : \mu_X^B(x) = 1\}, \\ X_{B^\oplus} &= \{x \in U : \mu_X^B(x) > 0\}, \\ X_{B^\pm} &= \{x \in U : 0 < \mu_X^B(x) < 1\}. \end{aligned}$$

In essence, classical rough sets are the weakest form of the quantitative idea above which abstracts away from the "degree to which" and simply uses "possibly in", instead.

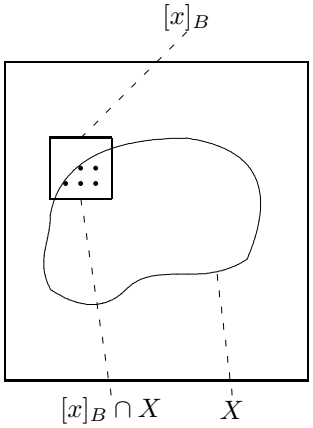

Figure 5.6: Rough membership function.

5.6 Decision Systems and Decision Rules

Rough set techniques are often used as a basis for supervised learning using tables of data (see Chapter 15). In many cases the target of a classification task is represented by an additional attribute called a decision attribute. Information systems of this kind are called decision systems.

Definition 5.6.1 Let $\langle U, A\rangle$ be an information system. A *decision system* is any system of the form $\mathcal{A} = \langle U, A, d\rangle$, where $d \notin A$ is the *decision attribute* and A is a set of *conditional attributes*, or simply *conditions*. ■

Let $\mathcal{A} = \langle U, A, d\rangle$ be given and let $V_d = \{v_1, \ldots, v_{r(d)}\}$. Decision d determines a partition $\{X_1, \ldots, X_{r(d)}\}$ of the universe U, where $X_k = \{x \in U \ : \ d(x) = v_k\}$ for $1 \leq k \leq r(d)$. The set X_i is called the i-th *decision class* of $\mathcal{A}$. By $X_{d(u)}$ we denote the decision class $\{x \in U \ : \ d(x) = d(u)\}$, for any $u \in U$.

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

One can generalize the above definition to the case of decision systems of the form $\mathcal{A} = \langle U, A, D\rangle$ where the set $D = \{d_1, ...d_k\}$ of decision attributes and A are assumed to be disjoint. Formally this system can be treated as the decision system $\mathcal{A} = \langle U, C, d_D\rangle$ where $d_D(x) = \langle d_1(x), ..., d_k(x)\rangle$ for $x \in U$.

Similarly as in the case of information systems, decision systems can be naturally represented as *decision tables*. The first column od a decision table contains objects from U, the next columns contain values of conditional attributes and the remaining columns contain decision attributes. In order to

separate conditional and decision attributes we use a double vertical line.

Example 5.6.2 Consider the situation described in Example 5.1.1. A similar representation of the road could be obtained using a more accurate camera by collecting data from the camera and creating an approximate description of the concept of road. In this example, a similar granularity of approximation dependent on the camera's resolution is assumed. Consider Figure 5.7. The decision table on the lefthand side of the figure contains some example data, where

- objects in the first column are points
- there are two attribute columns, *Pixel* and *Inside*.
  - *Pixel* refers to a pixel which is defined as an elementary set containing coordinates. For example, pixel (i, j) contains points $\{(x, y) : i - 1 \leq x < i, j - 1 \leq x < j\}$.
  - *Inside* represents information about whether a point is inside the boundaries of a road object.

The additional column "Approximation of Inside", provides us with the approximation obtained from the available data, where:

- "+" means that a given point is inside of the road because it is contained in a pixel which is totally included in the road;
- "±" means that it is unknown whether the point is inside of the road, because it is contained in a pixel, part of which is inside and part of which is outside of the road;
- "− means that the point is outside of the road because it is contained in a pixel which is totally outside of the road. ∎

Let $\mathcal{A}$ be a decision system. A *condition template* of $\mathcal{A}$ is any template containing conditional attributes only. A *decision rule* for $\mathcal{A}$ is any expression of the form $\varphi \Rightarrow \psi$, where φ is a condition template and ψ is a descriptor containing a decision attribute. Formulas φ and ψ are referred to as the *predecessor* and the *successor of the decision rule*. Decision rules are often called *if-then rules*.

A decision rule $\varphi \Rightarrow \psi$ is True in $\mathcal{A}$ iff $\|\varphi\|_{\mathcal{A}} \subseteq \|\psi\|_{\mathcal{A}}$. Otherwise, one can measure its *truth degree* by introducing some inclusion measure of $\|\varphi\|_{\mathcal{A}}$ in $\|\psi\|_{\mathcal{A}}$. For example, one such measure which is widely used, is called a *confidence coefficient* and is defined as,

$$\begin{cases} \frac{|\|\varphi \wedge \psi\|_{\mathcal{A}}|}{|\,\|\varphi\|_{\mathcal{A}}|} & \text{for } |\,\|\varphi\|_{\mathcal{A}}| \neq 0 \\ 1 & \text{otherwise.} \end{cases}$$

Point	*Pixel*	*Inside*	*Approximation of Inside*
$(4.60, 5.50)$	$(5, 6)$	True	$+$
$(4.50, 4.70)$	$(5, 5)$	True	$\pm$
$(4.90, 4.01)$	$(5, 5)$	False	$\pm$
$(4.50, 3.50)$	$(5, 4)$	False	$\pm$
$(4.01, 3.99)$	$(5, 4)$	True	$\pm$
$(5.50, 5.50)$	$(6, 6)$	True	$+$
$(5.50, 4.50)$	$(6, 5)$	True	$\pm$
$(5.50, 4.10)$	$(6, 5)$	False	$\pm$
$(5.50, 3.40)$	$(6, 4)$	False	$-$
$(6.40, 5.40)$	$(7, 6)$	True	$+$
$(6.50, 4.80)$	$(7, 5)$	True	$\pm$
$(6.50, 4.20)$	$(7, 5)$	False	$\pm$
$(6.30, 3.40)$	$(7, 4)$	False	$-$

Figure 5.7: Decision table for determining *Inside* of Example 5.6.2

Another measure of non-classical inclusion, called *support of the rule*, is defined as

$$\frac{|\,\|\varphi \wedge \psi\|_{\mathcal{A}}|}{|U|}.$$

An interesting class of decision rules consists of *minimal decision rules* in a given decision system $\mathcal{A}$, i.e., rules which are true in $\mathcal{A}$ but become not true after removing any conditional descriptor from them.

Each object x in a decision table determines a decision rule,

$$\bigwedge_{a \in C} (a = a(x)) \Rightarrow (d = d(x)),$$

where C is the set of conditional attributes and d is the decision attribute. Decision rules corresponding to some objects can have the same condition parts but different decision parts. Such rules are called *inconsistent* (*nondeterministic*, *conflicting*); otherwise the rules are referred to as *consistent* (*certain*, *deterministic*, *nonconflicting*) rules. Decision tables containing inconsistent decision rules are called *inconsistent* (*nondeterministic*, *conflicting*); otherwise the table is *consistent* (*deterministic*, *nonconflicting*).

Numerous methods based on the rough set approach combined with Boolean reasoning techniques have been developed for decision rule generation. When a set of rules have been induced from a decision table containing a set of training examples, they can be inspected to determine if they reveal any

novel relationships between attributes that are worth pursuing further. In addition, the rules can be applied to a set of unseen cases in order to estimate their classification power.

For a systematic overview of rule generation and application methods the reader is referred to Chapter 15 in this book which covers the topic of machine learning and knowledge discovery.

5.7 Inducing Consistent Concept Descriptions

Up to now, we considered the approximations of raw data represented as decision tables. It is often the case that, for a particular application domain, additional qualitative knowledge about dependencies between concepts are often available in the form of expert knowledge. In this section, we consider methods for inducing concept approximations constrained by this additional knowledge.

Consider the traffic scenario domain used in previous examples. Given a set of facts represented by decision tables, an example of a dependency between concept approximations could be the rule,

> if a road is slippery and the speed of a car is high, then there is a high chance that an accident involving the car will occur.

An interesting issue arises as to how concept approximations can be induced using both the raw data in decision tables and qualitative knowledge associated with it. There are a number of approaches which immediately come to mind. For example,

- one can develop strategies for directly generating decision rules which preserve the qualitative dependencies between approximated concepts;
- one can tune decision rules generated from raw data relative to the qualitative dependencies.

Yet another approach, and one we will pursue in this book, involves using some ideas from the area of nonmonotonic reasoning together with the second approach above. For example, consider a case where our decision tables contain the conditional attributes d_1, d_2 and the decision attribute d_3. The following dependency might be provided by a domain expert:

> if $d_1 =$ high is known and $d_2 =$ medium is consistent with the available information then $d_3 =$ dangerous by default.

Given a decision table containing these attributes, we would first generate the minimal decision rules representing the data[2]. We would then view the dependencies as representing constraints between the lower and upper approximations of the various decision classes. This can be done by interpreting the phrase "is known" as a lower approximation of its respective argument and the phrase "is consistent with the available information" as the upper approximation of its respective argument[3]. The minimal decision rules could then be tuned using the constraints between the lower and upper approximations of the decision classes in the qualitative dependency rules.

In the following, we provide an example of the main idea. The illustrated methodology is fully elaborated in the second and third part of this book.

Example 5.7.1 Consider again the situation described in Example 5.1.1. Suppose now that one wants to determine whether a given car is completely within the boundaries of a road. The suitable information can be obtained on the basis of the pixels available at a particular level of resolution. Car c1 is completely within the road's boundaries, since all pixels covering c1 are completely within the road's boundaries; car c3 is outside of the road's boundaries, since all pixels covering c3 are outside of the road's boundaries; and it is unknown whether car c2 is completely within or outside of the road's boundaries, since some of the pixels covering c2 are in the boundary region of the road.

However, in certain cases when additional knowledge is provided, one can classify c2 as being inside the road's boundaries. Namely, consider the following rule expressing knowledge about the road domain:

> if it is consistent with the available information that a car is completely within the boundaries of a road and its speed is known to be high then assume by default that the car is within the boundaries of the road.

The rule can be translated into the terminology of rough sets as follows:

> if a car is in the boundary region of a road and its speed is classified as high then assume that the car is completely within the boundaries of the road.

In the presence of additional information that car c2 is moving with high speed, one can conclude that it is completely within the boundaries of the road. ■

[2]For a technique of generating minimal rules see Chapter 15.

[3]Observe that rules of the form considered in the example are easily expressible in nonmonotonic formalisms, like default reasoning or circumscription (for more details see, e.g., Chapters 4 and 10).

5.8 Dependency of Attributes

An important issue in data analysis is to discover dependencies between attributes. Intuitively, a set of attributes D depends totally on a set of attributes C 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.

Formally, a dependency between attributes can be defined as follows.

Definition 5.8.1 Let $\mathcal{A} = \langle U, A \rangle$ be an information system, and let D and C be subsets of A. By a *positive region of a partition* U/D with respect to C, denoted by $POS_C(D)$, we mean the set[4]

$$POS_C(D) = \bigcup_{X \in U/D} X_{C^+}.$$

The *degree of the dependency* $C \mapsto D$, $\gamma_{\mathcal{A}}(C, D)$, is defined by

$$\gamma_{\mathcal{A}}(C, D) \stackrel{\text{def}}{=} \frac{|POS_C(D)|}{|U|}.$$

We say that D *depends on* C *to degree* k, denoted by $C \mapsto_k D$, if $k = \gamma_{\mathcal{A}}(C, D)$.

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. ∎

Observe that, for any $C, D \subseteq A$, $0 \leq \gamma_{\mathcal{A}}(C, D) \leq 1$.

If D depends totally on C then $IND_{\mathcal{A}}(C) \subseteq IND_{\mathcal{A}}(D)$. This means that the partition generated by C is finer than the partition generated by D. Notice, that the concept of dependency discussed above corresponds to that considered in relational databases.

A geometrical interpretation of the positive region $POS_C(D)$, is shown in Figure 5.8, where:

- dashed boxes represent a partition created by C
- the set of equivalence classes of $IND_{\mathcal{A}}(D)$, $U/D = \{X_1, X_2, X_3\}$, where the three partitions are separated by the thick black line
- the union of the double-lined areas represents $POS_C(D)$

[4]Recall that U/D denotes the set of equivalence classes of relation $IND_{\mathcal{A}}(D)$.

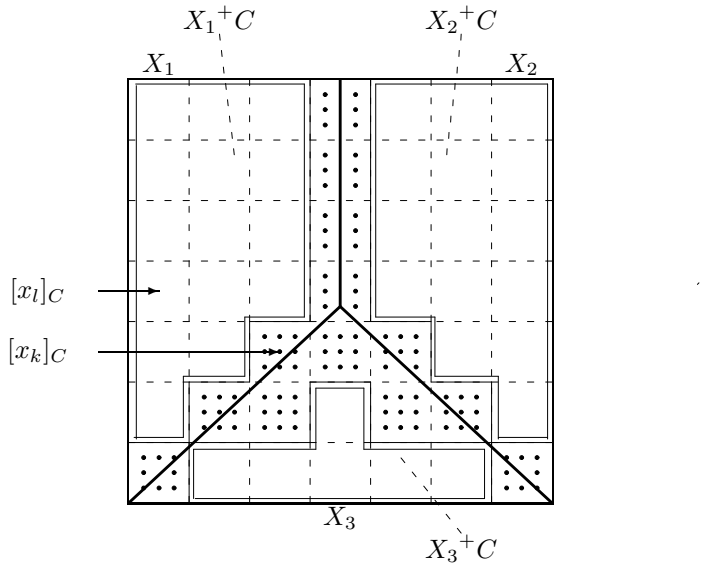

Figure 5.8: Geometrical interpretation of positive region $POS_C(D)$.

- the dotted boxes represent the boundary region between classes.

In this example, it is clear that there is only a degree of dependency between C and D because some partitions in C do not uniquely map to a partition in U/D. For instance, if $x \in [x_k]_C$, we are unable to uniquely determine whether $x \in X_1$ or $x \in X_3$. On the other hand, if $x \in [x_l]_C$ then $x \in X_1$.

In summary, 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.

5.9 Bibliographic Notes

In this chapter, we considered the basic concepts of rough set theory. The theory itself was originally proposed by Pawlak [117].

Several generalizations of the classical rough set approach have been reported in the literature (for references see the papers and bibliography in [109, 110, 83, 115, 125, 126, 121, 156, 149]).

The approach based on inclusion functions has been generalized to the rough mereological approach in [123, 147, 127]. A discussion of partial containment