

6.3 Rough Set Theory

A tool to deal with uncertainty, called rough set theory, was introduced by Z. Pawlak (1982). This theory is especially well suited to deal with inconsistencies in the process of knowledge acquisition. In the presented approach, inconsistencies are not corrected. The key issue is to compute lower and upper approximations of partitions, the fundamental concepts of rough set theory. On the basis of lower and upper approximations, two different sets of rules are computed: certain and possible. Certain rules are categorical and may be further employed using classical logic. Possible rules are supported by existing data, although conflicting data may exist as well. Possible rules may be processed further using either classical logic or any theory to deal with uncertainty.

Note that the presented approach may be combined with any other approach to uncertainty. An advantage of the method is that certain and possible rules are processed separately (i.e., two parallel inference engines may be used). Another

advantage of rough set theory is that it does not need any preliminary or additional information about data (like prior probability in probability theory, basic probability number in Dempster-Shafer theory, and grade of membership in fuzzy set theory).

6.3.1 Rough Sets

Let U be a nonempty set, called the *universe*, and let R be an equivalence relation on U , called an *indiscernibility relation*. An ordered pair $A = (U, R)$ is called an *approximation space*. For any element x of U , the equivalence class of R containing x is denoted $[x]_R$. Equivalence classes of R are called *elementary sets* in A . We assume that the empty set is also elementary.

Any finite union of elementary sets in A is called a *definable set* in A .

Let X be a subset of U . We wish to define X in terms of definable sets in A . Thus, we need two more concepts.

A *lower approximation* of X in A , denoted $\underline{R}X$, is the set

$$\{x \in U \mid [x]_R \subseteq X\}.$$

An *upper approximation* of X in A , denoted $\overline{R}X$, is the set

$$\{x \in U \mid [x]_R \cap X \neq \emptyset\}.$$

The lower approximation of X in A is the greatest definable set in A , contained in X . The upper approximation of X in A is the least definable set in A containing X . Time complexity of algorithms for computing lower and upper approximations of any set X is $O(m^2)$, where m is the cardinality of set U of entities.

For example, $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$, and R is an equivalence relation that induces the following partition:

$$R^* = \{\{x_1\}, \{x_2, x_3\}, \{x_4, x_5\}, \{x_6, x_7\}, \{x_8\}\},$$

that is, the equivalence classes of R are $\{x_1\}$, $\{x_2, x_3\}$, $\{x_4, x_5\}$, $\{x_6, x_7\}$, and $\{x_8\}$. Let X be equal to $\{x_1, x_2, x_3, x_5, x_7\}$. Then the lower approximation of X in $A = (U, R)$ is the set

$$\underline{R}X = \{x_1, x_2, x_3\},$$

and the upper approximation of X in A is the set

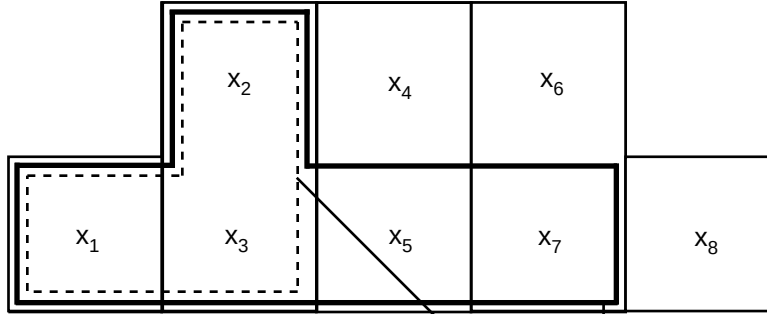
$$\overline{R}X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\},$$

(see Figure 6.12).

Let X and Y be subsets of U . Lower and upper approximations of X and Y in A have the following properties

$$\underline{R}X \subseteq X \subseteq \overline{R}X,$$

(a) Lower Approximation of X in A



(b) Upper Approximation of X in A

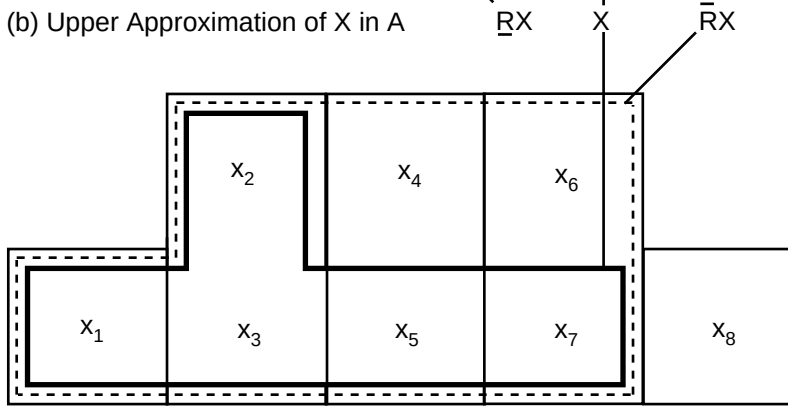


Figure 6.12 Sets X , $\underline{RX}$, and $\overline{RX}$

$$\underline{RU} = U = \overline{RU},$$

$$\underline{R}\emptyset = \emptyset = \overline{R}\emptyset,$$

$$\underline{R}(X \cup Y) \supseteq \underline{RX} \cup \underline{RY},$$

$$\overline{R}(X \cup Y) = \overline{RX} \cup \overline{RY},$$

$$\underline{R}(X \cap Y) = \underline{RX} \cap \underline{RY},$$

$$\overline{R}(X \cap Y) \subseteq \overline{RX} \cap \overline{RY},$$

$$\underline{R}(X - Y) \subseteq \underline{RX} - \underline{RY},$$

$$\overline{R}(X - Y) \supseteq \overline{RX} - \overline{RY},$$

$$\begin{aligned}
\underline{R}(-X) &= -\bar{R}X, \\
\bar{R}(-X) &= -\underline{R}X, \\
\underline{R}X \cup \bar{R}(-X) &= X, \\
\underline{R}(\underline{R}X) &= \bar{R}(\underline{R}X) = \underline{R}X, \\
\bar{R}(\bar{R}X) &= \underline{R}(\bar{R}X) = \bar{R}X,
\end{aligned}$$

where $-X$ denotes the complement $U - X$ of X .

A *rough set* in A is the family of all subsets of U having the same lower and upper approximations in A . In the example from Figure 6.12, the rough set is equal to

$$\begin{aligned}
&\{\{x_1, x_2, x_3, x_4, x_6\}, \{x_1, x_2, x_3, x_4, x_7\}, \\
&\{x_1, x_2, x_3, x_5, x_6\}, \{x_1, x_2, x_3, x_5, x_7\}\}.
\end{aligned}$$

Let x be in U . We say that x is *certainly in* X if and only if $x \in \underline{R}X$, and that x is *possibly in* X if and only if $x \in \bar{R}X$. Our terminology originates from the fact that we want to decide if x is in X on the basis of a definable set in A rather than on the basis of X . This means that we deal with $\underline{R}X$ and $\bar{R}X$ instead of X . Since $\underline{R}X \subseteq X \subseteq \bar{R}X$, if x is in $\underline{R}X$ it is certainly in X . On the other hand, if x is in $\bar{R}X$, it is possibly in X . In the example from Figure 6.12, x_3 is certainly in X , and therefore possibly in X , x_4 is possibly but not certainly in X , and x_8 is not possibly in X .

6.3.2 Rough Definability of a Set

A rough-set classification of a set of entities has been presented in Pawlak (1984). The main idea is to classify any such set by properties of its lower and upper approximations. The notation of Chapter 3 pertaining to decision tables is used in the sequel. Let Q be a set of all attributes and decisions. Let U be a set of all entities. For any nonempty subset P of Q , an ordered pair $(U, \sim_P)$ is an approximation space A . For the sake of convenience, for any $X \subseteq U$, the lower approximation of X in A and the upper approximation of X in A is called *P-lower approximation of* X and *P-upper approximation of* X , and is denoted $\underline{P}X$ and $\bar{P}X$, respectively.

An example of a decision table is presented in Table 6.17. Let $Q = \{a, b, c, d\}$ be the set of all attributes and decisions. Let P be the set of all attributes, i.e., $P = \{a, b, c\}$. The partition generated by P, P^* , is equal to

$$\{\{x_1\}, \{x_2, x_3\}, \{x_4, x_5\}, \{x_6, x_7\}, \{x_8\}\}.$$

Let X be a set $\{x_1, x_2, x_3, x_5, x_7\}$. Then

$$\underline{P}X = \{x_1, x_2, x_3\}$$

and

Table 6.17 A Decision Table

	Attributes			Decision
	a	b	c	d
x_1	0	3	0	0
x_2	0	4	1	0
x_3	0	4	1	0
x_4	1	4	1	1
x_5	1	4	1	0
x_6	2	4	1	1
x_7	2	4	1	0
x_8	2	5	2	1

$$\bar{P}X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}.$$

Any union of equivalence classes of $\sim_P$ is called a *P-definable set*. A set $X \subseteq U$ which is not *P-definable* is called *P-undefinable*.

The set X is called *roughly P-definable* if and only if $\underline{P}X \neq \emptyset$ and $\bar{P}X \neq U$.

The set X is called *internally P-undefinable* if and only if $\underline{P}X = \emptyset$ and $\bar{P}X \neq U$.

The set X is called *externally P-undefinable* if and only if $\underline{P}X \neq \emptyset$ and $\bar{P}X = U$.

The set X is called *totally P-undefinable* if and only if $\underline{P}X = \emptyset$ and $\bar{P}X = U$.

For an internally *P-undefinable* set X we can not say certainly that any $x \in U$ is a member of X . For an externally *P-undefinable* set X we can not exclude any element $x \in U$ being possibly a member of X .

Note that a set X may be *P-undefinable* and *roughly P-definable*. Yet another possibility is a set X that is *P-definable* and not *roughly P-definable* (e.g., the empty set $\emptyset$). If a set X is not *roughly P-definable*, then it is *internally*, *externally*, or *totally P-undefinable*. In the example from Table 6.17, set $\{x_1, x_2, x_3\}$ is *P-definable*, set X is *roughly P-definable*, set $\{x_3, x_5, x_7\}$ is *internally P-undefinable*, and set $\{x_1, x_3, x_5, x_7, x_8\}$ is *externally P-undefinable*. In the example a *P-undefinable* set does not exist.

6.3.3 Rough Measures of a Set

Measures of uncertainty based on rough set theory have auxiliary value only. In the rough-set approach, the set X is described by its lower and upper approximations.

A quality of lower approximation of X by P , denoted $\gamma_P(X)$, is equal to

$$\frac{|PX|}{|U|}.$$

In Pawlak (1984), $\gamma_P(X)$ is called quality of approximation of X by P .

Thus, the quality of lower approximation of X by P is the ratio of the number of all certainly classified entities by attributes from P as being in X to the number of all entities of the system. It is a kind of relative frequency. Note that $\gamma_P(X)$ is a belief function according to Dempster-Shafer theory (Shafer, 1976) because it satisfies all three conditions of the definition of belief function. There are essential connections between rough set theory and Dempster-Shafer theory. For example, lower and upper approximations of rough set theory exist under the names of inner and outer reductions, respectively, in the monograph by G. Shafer (1976). However, both theories have been developed separately and both have well-established records of publications. The main difference is in the emphasis: Dempster-Shafer theory uses belief function as a main tool, while rough set theory makes use of the rough set, i.e., the family of all sets with common lower and upper approximations.

Let $P^* = \{U_1, U_2, \dots, U_k\}$ be a partition of U generated by P . It is not difficult to recognize that the basic probability assignment m that corresponds to γ_P is given by

$$m(U_i) = \frac{|U_i|}{|U|},$$

where $i = 1, 2, \dots, k$, and $m(Y) = 0$ for all other $Y \subseteq U$. Therefore, focal elements (Shafer, 1976) of γ_P are elementary sets in the approximation space $(U, \mathcal{P}_P)$, and the core of γ_P is the set U .

A quality of upper approximation of X by P , denoted $\bar{\gamma}_P(X)$, is

$$\frac{|\bar{P}X|}{|U|}.$$

The quality of upper approximation of X by P is the ratio of the number of all possibly classified entities by attributes from P as being in X to the number of all entities of the system. Therefore, it is again a kind of relative frequency. The quality of upper approximation of X by P is a plausibility function from the Dempster-Shafer theory viewpoint (Shafer, 1987), called an upper probability function in Shafer (1976). In the example from Table 6.17,

$$\gamma_P(X) = \frac{3}{8} = 0.375, \text{ and } \bar{\gamma}_P(X) = \frac{7}{8} = 0.875.$$

6.3.4 Partitions

In this section, the concept of rough definability will be expanded to include partitions of the set of all entities. A partition is of great interest because in the process of learning from examples, rules are derived from partitions, generated by single decisions.

It is quite natural to expect that all concepts of rough definability of a set may be transformed into similar ones for partitions. Surprisingly, this is not true. For example, as follows from Grzymala-Busse (1988), externally undefinable partitions do not exist. Instead, they coalesce with totally undefinable partitions.

Let $\mathcal{X} = \{X_1, X_2, \dots, X_n\}$ be a partition of U . P -lower and P -upper approximations of $\mathcal{X}$ by P , denoted $\underline{P}\mathcal{X}$ and $\bar{P}\mathcal{X}$, respectively, are the following sets:

$$\underline{P}\mathcal{X} = \{\underline{P}X_1, \underline{P}X_2, \dots, \underline{P}X_n\}$$

and

$$\bar{P}\mathcal{X} = \{\bar{P}X_1, \bar{P}X_2, \dots, \bar{P}X_n\}.$$

The partition $\mathcal{X}$ of U is called P -definable if and only if $\underline{P}\mathcal{X} = \bar{P}\mathcal{X}$. The partition $\mathcal{X}$ of U that is not P -definable is called P -undefinable.

The partition $\mathcal{X}$ of U is called *roughly* P -definable, *weak* if and only if there exists a number $i \in \{1, 2, \dots, n\}$ such that $\underline{P}X_i \neq \emptyset$. For a roughly P -definable, weak partition $\mathcal{X}$ of U , there exists a number $j \in \{1, 2, \dots, n\}$ such that $\bar{P}X_j \neq U$ (Grzymala-Busse, 1988).

The partition $\mathcal{X}$ of U is called *roughly* P -definable, *strong* if and only if $\underline{P}X_i \neq \emptyset$ for each i from the set $\{1, 2, \dots, n\}$. In a roughly P -definable, strong partition $\mathcal{X}$ of U , $\bar{P}X_i \neq U$ for each i from the set $\{1, 2, \dots, n\}$ (Grzymala-Busse, 1988).

The partition $\mathcal{X}$ of U is called *internally* P -undefinable, *weak* if and only if $\underline{P}X_i = \emptyset$ for each $i \in \{1, 2, \dots, n\}$ and there exists $j \in \{1, 2, \dots, n\}$ such that $\bar{P}X_j \neq U$.

The partition $\mathcal{X}$ of U is called *internally* P -undefinable, *strong* if and only if $\underline{P}X_i = \emptyset$ and $\bar{P}X_i \neq U$ for each i from the set $\{1, 2, \dots, n\}$.

Finally, the partition $\mathcal{X}$ of U is called *totally* P -undefinable if and only if $\underline{P}X_i = \emptyset$ and $\bar{P}X_i = U$ for each $i \in \{1, 2, \dots, n\}$. For example, for the decision table presented in Table 6.18, let $P = \{a, b\}$ and $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$. Then

$$\underline{P}\{c\}^* = \bar{P}\{c\}^* = \{\{x_1, x_2, x_3, x_4, x_5\}, \{x_6, x_7, x_8\}\},$$

Table 6.18 A Decision Table

	<u>Attributes</u>		<u>D e c i s i o n s</u>					
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
x_1	0	0	0	0	0	0	0	0
x_2	0	0	0	0	0	1	1	1
x_3	0	0	0	0	0	2	1	1
x_4	0	1	0	0	0	0	0	0
x_5	0	1	0	1	1	1	2	1
x_6	1	1	1	1	1	0	1	0
x_7	1	1	1	2	1	1	2	0
x_8	1	1	1	2	1	2	2	1

$$\underline{P}\{d\}^* = \{\{x_1, x_2, x_3\}, \emptyset\}, \bar{P}\{d\}^* = \{\{x_1, x_2, x_3, x_4, x_5\}, \{x_4, x_5, x_6, x_7, x_8\}\},$$

$$\underline{P}\{e\}^* = \{\{x_1, x_2, x_3\}, \{x_6, x_7, x_8\}\}, \bar{P}\{e\} = \{\{x_1, x_2, x_3, x_4, x_5\}, \{x_4, x_5, x_6, x_7, x_8\}\},$$

$$\underline{P}\{f\}^* = \{\emptyset\}, \bar{P}\{f\}^* = \{U, \{x_1, x_2, x_3, x_6, x_7, x_8\}\},$$

$$\underline{P}\{g\}^* = \{\emptyset\}, \bar{P}\{g\}^* = \{\{x_1, x_2, x_3, x_4, x_5\}, \{x_1, x_2, x_3, x_6, x_7, x_8\}, \{x_4, x_5, x_6, x_7, x_8\}\},$$

$$\underline{P}\{h\}^* = \{\emptyset\}, \bar{P}\{h\}^* = \{U\}.$$

Thus, the partition $\{c\}^*$ is P -definable, $\{d\}^*$ is roughly P -definable, weak, $\{e\}^*$ is roughly P -definable, strong, $\{f\}^*$ is internally P -undefinable, weak, $\{g\}^*$ is internally P -definable, strong, and $\{h\}^*$ is totally P -undefinable.

A quality of lower approximation of $\mathcal{X}$ by P is equal to

$$\frac{\sum_{i=1}^n |\underline{P}X_i|}{|U|},$$

and a quality of upper approximation of $\mathcal{X}$ by P is equal to

$$\frac{\sum_{i=1}^n |\bar{P}X_i|}{|U|}.$$

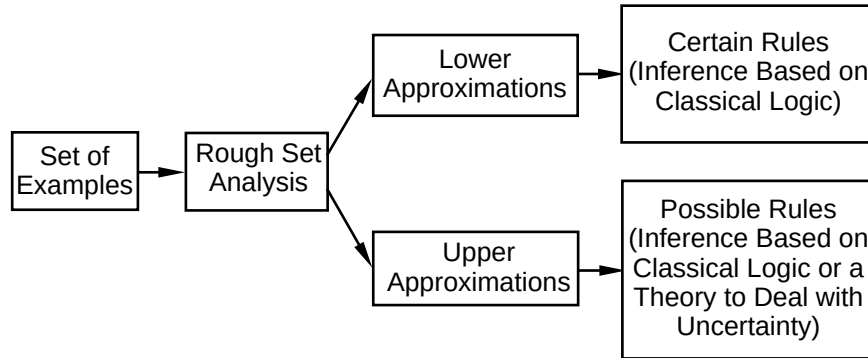


Figure 6.13 The Principle of the Use of Rough Set Theory for Learning from Examples

In the example from Table 6.18, the quality of lower approximation of $\{d\}^*$ is

$$\frac{3 + 0}{8} = 0.375,$$

and the quality of upper approximation of $\{d\}^*$ is

$$\frac{5 + 5}{8} = 1.25.$$

6.3.5 Certain and Possible Rules

The main idea of use of rough set theory for learning from examples is presented in Figure 6.13. A set of examples is given in the form of a decision table. All results of uncertainty are manifested finally by inconsistent information in the decision table.

Another example of a decision table is given by Table 6.19, which may be interpreted as containing data about patients denoted by numbers 0, 1,..., 9. Table 6.19 includes inconsistencies. For example, patients 3 and 4, described by the same values of attributes, are classified differently by a decision: Patient 3 is free from influenza, while patient 4 has it. Another pair of inconsistencies is introduced by patients 6 and 7. In the example, the partition of patients as being sick with influenza (or, according to the preceding terminology, partition generated by $\{\text{Influenza}\}$) is

$$\chi = \{\{0, 1, 3, 5, 6\}, \{2, 4, 7, 8, 9\}\}.$$

Table 6.19 A Decision Table

Attributes					Decision
	Temperature	Dry_cough	Headache	Muscle_pain	Influenza
0	normal	absent	absent	absent	absent
1	normal	absent	present	present	absent
2	subfebrile	absent	present	present	present
3	subfebrile	present	absent	absent	absent
4	subfebrile	present	absent	absent	present
5	high	absent	absent	absent	absent
6	high	present	absent	absent	absent
7	high	present	absent	absent	present
8	high	present	present	present	present
9	high	present	present	present	present

For the above example,

$$\begin{aligned}
 & \{\text{Temperature, Dry_cough, Headache, Muscle_pain}\}^* \\
 &= \{\{0\}, \{1\}, \{2\}, \{3, 4\}, \{5\}, \{6, 7\}, \{8, 9\}\} \\
 &= \{\text{Temperature, Dry_cough, Headache}\}^*.
 \end{aligned}$$

Thus, the attribute Muscle_pain is redundant as a symptom. The set $\{\text{Temperature, Dry_cough, Headache}\}$ will be denoted P . The lower approximation $\underline{P}\mathcal{X}$ is equal to

$$\{\{0, 1, 5\}, \{2, 8, 9\}\}$$

and the upper approximation $\overline{P}\mathcal{X}$ is equal to

$$\{\{0, 1, 3, 4, 5, 6, 7\}, \{2, 3, 4, 6, 7, 8, 9\}\}.$$

Certain rules follow from $\underline{P}\mathcal{X}$. Every element of $\underline{P}\mathcal{X}$ is P -definable; hence, it may be represented by rules using attributes of set P . Furthermore, set $\{0, 1, 5\}$, an element of $\underline{P}\mathcal{X}$, is described by the certain fact (Influenza, absent). Corresponding certain rules are

$$\begin{aligned}
 & (\text{Temperature, normal}) \wedge (\text{Dry_cough, absent}) \wedge (\text{Headache, absent}) \rightarrow \\
 & (\text{Influenza, absent}),
 \end{aligned}$$

$(\text{Temperature, normal}) \wedge (\text{Dry_cough, absent}) \wedge (\text{Headache, present}) \rightarrow (\text{Influenza, absent}),$

$(\text{Temperature, high}) \wedge (\text{Dry_cough, absent}) \wedge (\text{Headache, absent}) \rightarrow (\text{Influenza, absent}).$

The preceding certain rules, after simplification by dropping conditions, become

$(\text{Temperature, normal}) \rightarrow (\text{Influenza, absent}),$

$(\text{Temperature, high}) \wedge (\text{Dry_cough, absent}) \rightarrow (\text{Influenza, absent}).$

Another element of $\underline{P}\mathcal{X}$, set $\{2, 8, 9\}$, implies the following certain rule (after simplification):

$(\text{Temperature, not normal}) \wedge (\text{Headache, present}) \rightarrow (\text{Influenza, present}).$

Possible rules are generated by $\bar{P}\mathcal{X}$. Again, every element of $\bar{P}\mathcal{X}$ is P -definable. On the other hand, set $\{0, 1, 3, 4, 5, 6, 7\}$, an element of $\bar{P}\mathcal{X}$, is described by the possible fact (Influenza, absent). Thus, possible rules, after simplification, are

$(\text{Temperature, normal}) \rightarrow (\text{Influenza, absent}),$

$(\text{Headache, absent}) \rightarrow (\text{Influenza, absent}).$

Possible rules that follow from the other element of $\bar{P}\mathcal{X}$, after simplification, are

$(\text{Temperature, subfebrile}) \rightarrow (\text{Influenza, present}),$

$(\text{Dry_cough, present}) \rightarrow (\text{Influenza, present}).$