A Decision-Theoretic Rough Set Model

Yiyu Yao and Jingtao Yao
Department of Computer Science
University of Regina
Regina, Saskatchewan, Canada S4S 0A2
\{yyao,jtyao\}@cs.uregina.ca

Special Thanks to Professor Günther Ruhe and Dr. Jingzhou Li for providing this opportunity.
Introduction to Rough Sets

- Rough sets is a new mathematical theory for dealing with vagueness and uncertainty.
- The theory is different from, and complementary to, fuzzy sets. It is a generalization of standard set theory.
- The theory is a concrete model of granular computing (GrC).
- The theory has been successfully applied in many fields. For example, machine learning, data mining, data analysis, medicine, cognitive science, expert systems, and many more.
Introduction to Rough Sets

• Basic assumption:
  Objects are defined, represented, or characterized based on a finite number of attributes or properties.

• Implications:
  We cannot distinguish some objects.
  We can only observe, measure, or define a certain set of objects as a whole rather than many individuals.
  Only some subsets in the power set can be measured or defined.

• Type of uncertainty: The uncertainty comes from our inability to distinguish certain objects.
Introduction to Rough Sets

• Basic problem:
  How to represent undefinable subsets based on definable subsets?

• Solution:
  An undefinable subsets are approximately represented by two definable subsets, called lower and upper approximations.
• A motivating example: information table

<table>
<thead>
<tr>
<th>Object</th>
<th>Height</th>
<th>Hair</th>
<th>Eyes</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>$o_1$</td>
<td>short</td>
<td>blond</td>
<td>blue</td>
<td>+</td>
</tr>
<tr>
<td>$o_2$</td>
<td>short</td>
<td>blond</td>
<td>brown</td>
<td>-</td>
</tr>
<tr>
<td>$o_3$</td>
<td>tall</td>
<td>red</td>
<td>blue</td>
<td>+</td>
</tr>
<tr>
<td>$o_4$</td>
<td>tall</td>
<td>dark</td>
<td>blue</td>
<td>-</td>
</tr>
<tr>
<td>$o_5$</td>
<td>tall</td>
<td>dark</td>
<td>blue</td>
<td>-</td>
</tr>
<tr>
<td>$o_6$</td>
<td>tall</td>
<td>blond</td>
<td>blue</td>
<td>+</td>
</tr>
<tr>
<td>$o_7$</td>
<td>tall</td>
<td>dark</td>
<td>brown</td>
<td>-</td>
</tr>
<tr>
<td>$o_8$</td>
<td>short</td>
<td>blond</td>
<td>brown</td>
<td>-</td>
</tr>
</tbody>
</table>
Introduction to Rough Sets

• Objects are described by three attributes: Height, Hair, and Eyes.

• If only the attribute Height is used, we obtain a partition:
  \[ \{ \{ o_1, o_2, o_8 \}, \{ o_3, o_4, o_5, o_6, o_7 \} \} . \]
  Based on the Height, we cannot distinguish objects \( o_1 \), \( o_2 \) and \( o_8 \). They represent the set of short people.

• If two attributes Height and Hair are used, we have the partition:
  \[ \{ \{ o_1, o_2, o_8 \}, \{ o_3 \}, \{ o_4, o_5, o_7 \}, \{ o_6 \} \} . \]
Introduction to Rough Sets

• For Height and Hair, the set of all definable subsets are:
  the empty set $\emptyset$.
  the entire set $U = \{o_1, \ldots, o_8\}$.
  the union of some blocks in the partition
  $\{\{o_1, o_2, o_8\}, \{o_3\}, \{o_4, o_5, o_7\}, \{o_6\}\}$.
  For example, $\{o_1, o_2, o_3, o_8\}$ is a definable set.

• The set $+ = \{o_1, o_3, o_6\}$ is a undefinable subset.

• $+$ can be approximated by two definable subsets from below and above:

$$\{o_3\} \subseteq \{o_1, o_3, o_6\} \subseteq \{o_1, o_2, o_3, o_6, o_8\}.$$
Significance of Rough Set Theory

• It provides a formal theory for dealing with a particular type of uncertainty induced by indistinguishability.
• It precisely defines the notion of what is definable and what is undefinable.
• It presents a philosophy of representing what is undefinable (unknown) based on what is definable (known).
Rough Set Theory: Formal Development

• Let $U$ be a finite set called universe.
• Let $E$ be an equivalence relation on $U$, that is, $E$ is reflexive, symmetric and transitive.
• Let $U/E$ be the partition induced by the equivalence relation.
• Let $[x]_E$ denote the equivalence class contain $x$.
• The pair $apr = (U, E)$ is called an approximation space.
• Let $\text{Def}(U)$ be the family of all definable subsets.
Rough Set Theory: Granule based definition

For any subset $X \subseteq U$, a pair of lower and upper approximations is defined by:

- The lower approximation $\underline{apr}(X)$ is the union of equivalence classes that are subsets of $X$.

- The upper approximation $\overline{apr}(X)$ is the union of equivalence classes that have non-empty overlap with $X$.

- Granule based definition provides a model for granular computing.
Rough Set Theory: Element based definition

For any subset \( X \subseteq U \), a pair of lower and upper approximations is defined by:

\[
\underline{apr}(X) = \{ x \mid \forall y \in U [xEy \implies y \in X] \},
\]
\[
\overline{apr}(X) = \{ x \mid \exists y \in U [xEy \land y \in X] \}.
\]

- An element \( x \) belongs to the lower approximation \( \underline{apr}(X) \) if all its equivalent elements belong to \( X \).
- An element \( x \) belongs to the upper approximation \( \overline{apr}(X) \) if at least one of its equivalent elements belongs to \( X \).
- Element based definition relates rough set theory to modal logic.
Rough Set Theory: Sub-system based definition

For any subset $X \subseteq U$, a pair of lower and upper approximations is defined by:

$$\underline{apr}(X) = \bigcup\{Y \mid Y \in \text{Def}(U) \land Y \subseteq X\},$$

$$\overline{apr}(X) = \bigcap\{Y \mid Y \in \text{Def}(U) \land X \subseteq Y\}.$$

- The lower approximation $\underline{apr}(X)$ is the largest definable set contained in $X$.
- The upper approximation $\overline{apr}(X)$ is the smallest definable set containing $X$.
- It is related to topological space, closure systems, and other mathematical systems, as well as belief functions.
Rough Set Theory: Algebraic systems

• The pair of approximations can be viewed as a pair of dual unary set-theoretic operators called approximation operators.

• The system \((2^U, \neg, \text{apr}, \overline{\text{apr}}, \cap, \cup)\) is an extension of standard set algebra \((2^U, \neg, \cap, \cup)\) with two added unary operators. It is called a rough set algebra.

• The rough set algebra is an example of Boolean algebras with added operators.
Rough Set Theory: Rough Classification

- Based on the lower and upper approximations, the universe $U$ can be divided into three disjoint regions, the positive region $\text{POS}(X)$, the negative region $\text{NEG}(X)$, and the boundary region $\text{BND}(X)$:
  \[
  \text{POS}(X) = \overline{\text{apr}}(X),
  \]
  \[
  \text{NEG}(X) = U - \overline{\text{apr}}(X),
  \]
  \[
  \text{BND}(X) = \text{apr}(X) - \overline{\text{apr}}(X).
  \]

- The boundary region consists of objects that cannot be classified without uncertainty, due to our inability to differentiate some different objects.
Rough Classification: Another formulation

- For $X \subseteq U$, its rough membership function is defined by:

$$\mu_X(x) = \frac{|[x]_E \cap X|}{|[x]_E|} = P(X \mid [x]_E).$$

- The three regions are then defined by:

$$\text{POS}(X) = \{x \in U \mid \mu_X(x) = 1\},$$
$$\text{NEG}(X) = \{x \in U \mid \mu_X(x) = 0\},$$
$$\text{BND}(X) = \{x \in U \mid 0 < \mu_X(x) < 1\}.$$

- Obviously, they use extreme values of $\mu_X$, i.e., 0 and 1.
Rough Classification: Observations

- All elements with non-zero and non-full membership values will be classified into boundary region.

- The quantitative information given by the conditional probability $P(X \mid [x]_E)$ is not taken into consideration.

- In practice, a not so rigid classification may be more useful. An object may be classified into the possible region if the conditional probability is sufficiently large. Likewise, an object may be classified into the negative region if the conditional probability is sufficiently small.
Decision-Theoretic Rough Sets:

- Basic question:
  How do we determine the threshold values for deciding the three regions?

- Answer:
  Use the Bayesian decision procedure.
Bayesian decision procedure

- Let $\Omega = \{w_1, \ldots, w_s\}$ be a finite set of $s$ states.
- Let $\mathcal{A} = \{a_1, \ldots, a_m\}$ be a finite set of $m$ possible actions.
- Let $P(w_j|\mathbf{x})$ be the conditional probability of an object $x$ being in state $w_j$ given that the object is described by $\mathbf{x}$.
- Let $\lambda(a_i|w_j)$ denote the loss, or cost, for taking action $a_i$ when the state is $w_j$. 

18
Bayesian decision procedure

- For an object with description $x$, suppose action $a_i$ is taken.

- Since $P(w_j|x)$ is the probability that the true state is $w_j$ given $x$, the expected loss associated with taking action $a_i$ is given by:

$$R(a_i|x) = \sum_{j=1}^{s} \lambda(a_i|w_j)P(w_j|x).$$

The quantity $R(a_i|x)$ is also called the conditional risk.

- Given description $x$, a decision rule is a function $\tau(x)$ that specifies which action to take.
• The overall risk is defined by:

\[ R = \sum_{x} R(\tau(x)|x)P(x), \]

where the summation is over the set of all possible descriptions of objects.

• Bayesian decision procedure: For every \( x \), compute the conditional risk \( R(a_i|x) \) for \( i = 1, \ldots, m \) and then select the action for which the conditional risk is minimum. If more than one action minimizes \( R(a_i|x) \), any tie-breaking rule can be used.
A Simple Example

• Set of states:
  \( s_0 \) – the meeting will be over in less than 2 hours,
  \( s_1 \) – the meeting will be more than 2 hours.

• \( P(s_0) = 0.8, \quad P(s_1) = 0.2. \)

• Set of actions:
  \( a_0 \) – put the car on meter (pay $2.00),
  \( a_1 \) – put the car in the parking lot (pay $7.00).

• Loss function:

  \[ \lambda(a_0 \mid s_0) = u(\$2.00) = 2, \quad \lambda(a_1 \mid s_0) = u(\$7.00) = 10, \]
  \[ \lambda(a_0 \mid s_1) = u(\$2.00 + \$15.00) = 60, \quad \lambda(a_1 \mid s_1) = u(\$7.00) = 10. \]
• The expected cost of actions $a_0$ and $a_1$:

\[
R(a_0) = \lambda(a_0 \mid s_0)P(s_0) + \lambda(a_0 \mid s_1)P(s_1)
= 2 \times 0.8 + 60 \times 0.2 = 13.6,
\]

\[
R(a_1) = \lambda(a_1 \mid s_0)P(s_0) + \lambda(a_1 \mid s_1)P(s_1)
= 10 \times 0.8 + 10 \times 0.2 = 10.0.
\]

• Choose action $a_1$. 

22
Decision-Theoretic Rough Sets

• The set of states: $\Omega = \{X, \neg X\}$.

• The set of actions: $\mathcal{A} = \{a_1, a_2, a_3\}$, deciding POS($\mathcal{A}$),
  deciding NEG($\mathcal{A}$), and deciding BND($\mathcal{A}$), respectively.

• Description of $x$: $[x]_E$.

• Conditional probability:
  $P(X | [x]_E)$ and $P(\neg X | [x]_E) = 1 - P(X | [x]_E)$.

• Loss function: $\lambda_{i1} = \lambda(a_i | X)$, $\lambda_{i2} = \lambda(a_i | \neg X)$, and $i = 1, 2, 3$. 
• expected loss $R(a_i|[x]_E)$ associated with taking the individual actions can be expressed as:

$$R(a_1|[x]_E) = \lambda_{11} P(X|[x]_E) + \lambda_{12} P(\neg X|[x]_E),$$

$$R(a_2|[x]_E) = \lambda_{21} P(X|[x]_E) + \lambda_{22} P(\neg X|[x]_E),$$

$$R(a_3|[x]_E) = \lambda_{31} P(X|[x]_E) + \lambda_{32} P(\neg X|[x]_E).$$
• The Bayesian decision procedure leads to the following minimum-risk decision rules:

(P) If $R(a_1|x_E) \leq R(a_2|x_E)$ and $R(a_1|x_E) \leq R(a_3|x_E)$, decide POS($X$);

(N) If $R(a_2|x_E) \leq R(a_1|x_E)$ and $R(a_2|x_E) \leq R(a_3|x_E)$, decide NEG($X$);

(B) If $R(a_3|x_E) \leq R(a_1|x_E)$ and $R(a_3|x_E) \leq R(a_2|x_E)$, decide BND($X$).

• Based on $P(A|x_E) + P(\neg A|x_E) = 1$, the decision rules can be simplified by using only the probabilities $P(X|x_E)$. 
• Consider a special kind of loss functions with $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$.

• We have:

(P) If $P(X|[x]_E) \geq \gamma$ and $P(X|[x]_E) \geq \alpha$, decide POS($X$);

(N) If $P(X|[x]_E) \leq \beta$ and $P(X|[x]_E) \leq \gamma$, decide NEG($X$);

(B) If $\beta \leq P(X|[x]_E) \leq \alpha$, decide BND($X$);

$$\alpha = \frac{\lambda_{12} - \lambda_{32}}{(\lambda_{31} - \lambda_{32}) - (\lambda_{11} - \lambda_{12})},$$

$$\gamma = \frac{\lambda_{12} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{11} - \lambda_{12})},$$

$$\beta = \frac{\lambda_{32} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{31} - \lambda_{32})}.$$  (1)
Suppose further:
\[(\lambda_{12} - \lambda_{32})(\lambda_{21} - \lambda_{31}) \geq (\lambda_{31} - \lambda_{11})(\lambda_{32} - \lambda_{22}),\]
we have: \(\alpha \geq \gamma \geq \beta\).

This leads to the following decision rules:

(P1) If \(P(\mathbf{X} | [x]_E) \geq \alpha\), decide \(\text{POS}(\mathbf{X})\);

(N1) If \(P(\mathbf{X} | [x]_E) \leq \beta\), decide \(\text{NEG}(\mathbf{X})\);

(B1) If \(\beta < P(\mathbf{X} | [x]_E) < \alpha\), decide \(\text{BND}(\mathbf{X})\).
• When $\alpha = \beta$, we have $\alpha = \gamma = \beta$. In this case, we use the decision rules:

(P2) If $P(X|[x]_E) > \alpha$, decide POS($X$);

(N2) If $P(X|[x]_E) < \alpha$, decide NEG($X$);

(B2) If $P(X|[x]_E) = \alpha$, decide BND($X$).
• Example 1:
• $\lambda_{12} = \lambda_{21} = 1$, $\lambda_{11} = \lambda_{22} = \lambda_{31} = \lambda_{32} = 0$.
• $\alpha = 1 > \beta = 0$, $\alpha = 1 - \beta$, and $\gamma = 0.5$.
• classical rough sets:

\[
\begin{align*}
\text{POS}(X) &= \{x \in U \mid \mu_X(x) = 1\}, \\
\text{NEG}(X) &= \{x \in U \mid \mu_X(x) = 0\}, \\
\text{BND}(X) &= \{x \in U \mid 0 < \mu_X(x) < 1\}.
\end{align*}
\]
• Example 2:

• $\lambda_{12} = \lambda_{21} = 1$, $\lambda_{31} = \lambda_{32} = 0.5$, $\lambda_{11} = \lambda_{22} = 0$.

• $\alpha = \beta = \gamma = 0.5$.

• 0.5-classification:

\[
\begin{align*}
\text{POS}(X) &= \{ x \in U \mid \mu_X(x) > 0.5 \}, \\
\text{NEG}(X) &= \{ x \in U \mid \mu_X(x) < 0.5 \}, \\
\text{BND}(X) &= \{ x \in U \mid \mu_X(x) = 0.5 \}.
\end{align*}
\]
• Example 3:
• $\lambda_{12} = \lambda_{21} = 4, \quad \lambda_{31} = \lambda_{32} = 1, \quad \lambda_{11} = \lambda_{22} = 0$.
• $\alpha = 0.75, \beta = 0.25$ and $\gamma = 0.5$.
• $1/4$-classification:

$$\text{POS}(X) = \{ x \in U \mid \mu_X(x) \geq 0.75 \},$$
$$\text{NEG}(X) = \{ x \in U \mid \mu_X(x) \leq 0.25 \},$$
$$\text{BND}(X) = \{ x \in U \mid 0.25 < \mu_X(x) < 0.75 \}.$$
Conclusions

- The rough set theory provides a sound and useful framework to study many issues.
- The language of rough sets can be used to describe concisely many problems.
- The rough set theory has a solid foundation and is related to many other theories.
- The decision-theoretic rough set theory is a probabilistic generalization of standard rough set theory.
- The decision-theoretic rough set theory extends the application domain of rough sets.
Thank you!