Fuzzy and Rough Sets
Part II

Decision Systems Group
Brigham and Women’s Hospital,
Harvard Medical School

HST 951 Spring 2003
Harvard-MIT Division of Health Sciences and Technology
Overview

• Fuzzy sets
• Fuzzy logic and rules
• Rough sets and rules
• An example of a method for mining rough/fuzzy rules
• Uncertainty revisited
Crisp Sets

• A set with a characteristic function is called *crisp*

• Crisp sets are used to formally characterize a *concept*, e.g., even numbers

• Crisp sets have clear cut boundaries, hence do not reflect uncertainty about membership
Fuzzy Sets

• Zadeh (1965) introduced “Fuzzy Sets” where he replaced the characteristic function with membership

• $\chi_S : U \rightarrow \{0,1\}$ is replaced by $m_S : U \rightarrow [0,1]$

• Membership is a generalization of characteristic function and gives a “degree of membership”

• Successful applications in control theoretic settings (appliances, gearbox)
Fuzzy Sets

- Example: Let S be the set of people of normal height
- Normality is clearly not a crisp concept
Crisp Characterizations of Fuzzy Sets

- Support in $U$
  \[
  \text{Support}_U(S) = \{x \in U \mid m_S(x) > 0\}
  \]

- Containment
  \[
  A \subseteq B \text{ if and only if } m_A(x) \leq m_B(x) \text{ for all } x \in U
  \]

- There are non-crisp versions of the above
Fuzzy Set Operations

• Union
  \[ m_{A \cup B}(x) = \max(m_A(x), m_B(x)) \]

• Intersection
  \[ m_{A \cap B}(x) = \min(m_A(x), m_B(x)) \]

• Complementation
  \[ m_{U-A}(x) = 1 - m_A(x) \]

• Note that other definitions exist too
Fuzzy Memberships
Example

$m_A(x)$ $m_B(x)$
Fuzzy Union Example

$m_{A\cup B}(x)$

$m_A(x)$

$m_B(x)$
Fuzzy Intersection Example
Fuzzy Complementation Example

\[ m_{U-A}(x) \]

\[ m_A(x) \]

\[ m_{U-A}(x) \]
Fuzzy Relations

• The fuzzy relation $R$ between Sets $X$ and $Y$ is a fuzzy set in the Cartesian product $X \times Y$
• $m_R : X \times Y \rightarrow [0,1]$ gives the degree to which $x$ and $y$ are related to each other in $R$. 
Composition of Relations

• Two fuzzy relations $R$ in $X \times Y$ and $S$ in $Y \times Z$ can be composed into $R \circ S$ in $X \times Z$ as

$$m_{R \circ S}(x,z) = \max_{y \in Y} \left[ \min[m_R(x,y), m_S(y,z)] \right]$$
Composition Example
Probabilities of Fuzzy Events

• “Probability of cold weather tomorrow”

• $U = \{x_1, x_2, \ldots, x_n\}$, $p$ is a probability density, $A$ is a fuzzy set (event) in $U$

$$P(A) = \sum_{i=1}^{n} m_A(x_i) p(x_i)$$
Defuzzyfication

- Finding a single representative for a fuzzy set $A$ in $U = \{x_i|i \text{ in } \{1,\ldots,n\}\}$
- Max: $x$ in $U$ such that $m_A(x)$ is maximal
- Center of gravity:

$$\frac{\sum_{i=1}^{n} x_i m_A(x_i)}{\sum_{i=1}^{n} m_A(x_i)}$$
Alpha Cuts

• \( A \) is a fuzzy set in \( U \)
• \( A_\alpha = \{ x \mid m_A(x) \geq \alpha \} \) is the \( \alpha \)-cut of \( A \) in \( U \)
• Strong \( \alpha \)-cut is
  \( A_\alpha = \{ x \mid m_A(x) > \alpha \} \)
• Alpha cuts are crisp sets
Fuzzy Logic

• Different views
  – Foundation for reasoning based on uncertain statements
  – Foundation for reasoning based on uncertain statements where fuzzy set theoretic tools are used (original Zadeh)
  – As a multivalued logic with operations chosen in a special way that has some fuzzy interpretation
Fuzzy Logic

• Generalization of proposition over a set
• Let $\chi_S: U \rightarrow \{0,1\}$ denote the characteristic function of the set $S$
• Recall that in “crisp” logic
  $I(p(x)) = p(x) = \chi_{T(p)}(x)$
  where $p$ is a proposition and $T(p)$ is the corresponding truth set
Fuzzy Logic

• We extend the proposition
  \[ p: U \rightarrow \{0,1\} \]
to be a fuzzy membership
  \[ p: U \rightarrow [0,1] \]
• The fuzzy set associated with \( p \) corresponds to the truth set \( T(p) \) and \( p(x) \) is the degree of truth of \( p \) for \( x \)
• We extend the interpretation of logical formulae analogously to the crisp case
Fuzzy Logic Semantics

• Basic operations:
  - \( I(p(x)) = p(x) \)
  - \( I(\alpha \lor \beta) = \max(I(\alpha), I(\beta)) \)
  - \( I(\alpha \land \beta) = \min(I(\alpha), I(\beta)) \)
  - \( I(\neg \alpha) = 1 - I(\alpha) \)
Fuzzy Logic Semantics

• Implication:
  – Kleene-Dienes
    \[ I(\alpha \rightarrow \beta) = \max(I(\neg \alpha), I(\beta)) \]
  – Zadeh
    \[ I(\alpha \rightarrow \beta) = \max(I(\neg \alpha), \min(I(\alpha), I(\beta))) \]

• Dubois and Prade (1992) analyze other definitions of Implications
  – Zadeh
    \[ I(\alpha \rightarrow \beta) = \max(I(\neg \alpha), \min(I(\alpha), I(\beta))) \]

HST 951 Spring 2003
Fuzzy Rules

• “If x in A then y in B” is a relation R between A and B

• Two model types
  – Implicative: \((x \in A \rightarrow y \in B)\) is an upper bound
  – Conjunctive: \((x \in A \land y \in B)\) is a lower bound
  – Crisp motivation:
    \[ \chi_A(x) \land \chi_B(y) \leq \chi_R(x,y) \leq (1 - \chi_A(x)) \lor \chi_B(y) \]
Conjunctive Rule application

- $R: U \times U \rightarrow [0,1]$ is a rule
  If $p(x)$ then $q(y)$
- Using a generalized Modus Ponens
  $A'$
  $A \rightarrow B$
  $B'$
  \[ B'(y) = \max_x [\min[A'(x), R(x,y)]] \]
Rough Sets

• Pawlak 1982
• Approximation of sets using a collection of sets.
• Related to fuzzy sets (Zadeh 1965), in that both can be viewed as representations of uncertainty regarding set membership.
Rough Set: Set Approximation

\[ C_2 \quad C_3 \quad C_4 \]
Rough Set: Set Approximation
Rough Set: Set Approximation

- Approximation of $D$ by $\{C_1, C_2, C_3, C_4\}$:
  - $C_1$ definitely outside
  - $C_3$ definitely inside: lower approximation
  - $C_2 \cup C_4$ are boundary
  - $C_2 \cup C_3 \cup C_4$ are upper approximation
Rough Set: Set Approximation

- Given a collection of sets $C = \{C_1, C_2, C_3, \ldots\}$ and a set $D$, we define:
  
  - **Lower approximation** of $D$ by $C$,
    \[ D^L = \bigcup C_i \text{ such that } C_i \cap D = C_i \]
  
  - **Upper approximation** of $D$ by $C$,
    \[ D^U = \bigcup C_i \text{ such that } C_i \cap D \neq \emptyset \]
  
  - **Boundary** of $D$ by $C$,
    \[ D^U_L = D^U - D^L \]
Rough Set: Definition

- A set D is *rough* with respect to a collection of sets C if it has a non-empty boundary when approximated by C. Otherwise it is *crisp*.
Rough Set: Information System

- Universe U of elements, e.g., patients.
- Set A of features (attributes), functions f from U to some set of values V_f.
- (U,A) – information system

<table>
<thead>
<tr>
<th>Object no.</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ U = \{1,2,3,4,5,6,7,8,9\} \]
\[ A = \{a,b,c,d\} \]
\[ V_a = V_b = V_c = V_d = \{0,1\} \]
Rough Sets: Partition of $U$

- $E = \{(i,j) \in U \times U \mid abc(i) = abc(j)\}$, equivalence relation on $U$
  - $E(1) = \{1\} = C_1$
  - $E(2) = E(3) = E(4) = \{2,3,4\} = C_2$
  - $E(5) = E(6) = \{5,6\} = C_3$
  - $E(7) = E(8) = E(9) = \{7,8,9\} = C_4$

<table>
<thead>
<tr>
<th>Object no.</th>
<th>abc</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0,0,1)</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>(0,1,1)</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>(0,1,1)</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>(0,1,1)</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>(1,0,0)</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>(1,0,0)</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>(1,1,0)</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>(1,1,0)</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>(1,1,0)</td>
<td>0</td>
</tr>
</tbody>
</table>
Rough Sets: Approximating $D$

$D^U = \{2,3,4,5,6,7,8,9\} = C_2 \cup C_3 \cup C_4$

$D_L = \{5,6\} = C_3$

$D^U - D_L = \{2,3,4,7,8,9\} = C_2 \cup C_4$

HST 951 Spring 2003
Rough Sets: Approximate membership $\delta$

$$\delta(i) = \frac{|D \cap E(i)|}{|E(i)|}$$

- $\delta(1) = 0$
- $\delta(2) = \delta(3) = \delta(4) = 1/3$
- $\delta(5) = \delta(6) = 1$
- $\delta(7) = \delta(8) = \delta(9) = 2/3$
Rough Sets: Data Compression

Information: Partition given by equivalence. Find minimal sets of features that preserve information in table.

<table>
<thead>
<tr>
<th>Object no.</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Object no.</th>
<th>a</th>
<th>b</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Rough Sets: Discernibility Matrix

- $M_A = \{m_{ij}\}, A = \{a,b,c\}$
- $m_{ij} = \{a \in A \mid a(k) \neq a(l), k \in C_i, l \in C_j\}$

\[
M_A = \begin{array}{cccc}
{} & \{b\} & \{a,c\} & \{a,b,c\} \\
\{b\} & {} & \{a,b,c\} & \{a,c\} \\
\{a,c\} & \{a,b,c\} & {} & \{b\} \\
\{a,b,c\} & \{a,c\} & \{b\} & {}
\end{array}
\]

$C = \{\{b\},\{a,c\}\{a,b,c\}\}$ – set of non-empty entries of $M_A$
Minimal sets that have non-empty intersection with all elements of $C$ are $\{a,b\}$ and $\{b,c\}$ (Finding: Combinatorial)
These are called reducts of $(U,A)$
A reduct is a minimal set of features that preserves the partition.
Rough Sets: Extending $\delta$

- Problem: we only have the $\delta$ value for 4 of 8 possible input values. What is $\delta(1,1,1)$?
- By using compressed data that preserves the partition, we cover more of the feature space. All of it in this case. $\delta(1,1,1) = \delta(1,1) = 2/3$. 

<table>
<thead>
<tr>
<th>abc</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,0,1)</td>
<td>0</td>
</tr>
<tr>
<td>(0,1,1)</td>
<td>1/3</td>
</tr>
<tr>
<td>(1,0,0)</td>
<td>1</td>
</tr>
<tr>
<td>(1,1,0)</td>
<td>2/3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ab</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,0)</td>
<td>0</td>
</tr>
<tr>
<td>(0,1)</td>
<td>1/3</td>
</tr>
<tr>
<td>(1,0)</td>
<td>1</td>
</tr>
<tr>
<td>(1,1)</td>
<td>2/3</td>
</tr>
</tbody>
</table>
Rough Sets: Extending $\delta$

- Problem: extension not unique (and can extend to different parts of feature space).
- $\delta(1,1,1) = \delta(1,1) = 1/3$.
- Possible solution: generate several extensions and combine by voting. Generating all extensions is combinatorial.
- $\delta(1,1,1) = (2/3 + 1/3)/2 = 1/2$
Rough Sets: Classification rules

<table>
<thead>
<tr>
<th>Object no.</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Rules with right hand side support numbers:

\[
\begin{align*}
    & a(0) \ \text{AND} \ b(0) \Rightarrow d(0) & (1) \\
    & a(0) \ \text{AND} \ b(1) \Rightarrow d(1) \ \text{OR} \ d(0) & (1, 2) \\
    & a(1) \ \text{AND} \ b(0) \Rightarrow d(1) & (2) \\
    & a(1) \ \text{AND} \ b(1) \Rightarrow d(1) \ \text{OR} \ d(0) & (2, 1)
\end{align*}
\]
A Proposal for Mining Fuzzy Rules

- Recipe:
  1. Create rough information system by fuzzy discretization of data
  2. Compute rough decision rules
  3. Interpret rules as fuzzy rules
Fuzzy Discretization

- $A_1, A_2, \ldots, A_n$ are fuzzy sets in $U$
- $\text{disc}: U \rightarrow \{1,2,\ldots,n\}$
  \[ \text{disc}(x) = \{i \mid m_{A_i}(x) = \max\{m_{A_j}(x) \mid j \in \{1,2,\ldots,n\}\} \} \]
- $\text{disc}$ selects the index of the fuzzy set that yields the maximal membership
- Information system: subject each attribute value to $\text{disc}$
Fuzzy Rough Rules: Example

$A_1(3.14) = 0.6$
$A_1(0.1) = 0.3$
$A_2(3.14) = 0.5$
$A_2(0.1) = 0.8$

Object no. | a   | d  \\
------------|-----|-----
1           | 3.14| 0   \\
2           | 0.1 | 1   \\

Object no. | a   | d  \\
------------|-----|-----
1           | 1   | 0   \\
2           | 2   | 1   \\

if $A_1$ then $d=0$
if $A_2$ then $d=1$
Uncertainty

• Fuzzy sets can be said to model inherent vagueness
  Bob is "tall" - vagueness in the meaning of "tall", not in Bob's height

• Rough sets can be said to model ambiguity due to lack of information
And...

- Thank you for your attention