Data Mining Using Rough Sets

C.-C. Chan
Department of Computer Science
University of Akron
Akron, OH 44325-4003
USA
chan@uakron.edu
Outline

- Overview of Data Mining
- Basic Concepts of Rough Sets
- Fuzzy Sets, Dempster-Shafer’s Theory and Bayes’ Theorem
- Software Tools
- LERS (Learning from Examples using Rough Sets)
- A Rule-Based System for Data Mining
- Concluding Remarks
Data Mining (KDD)

- From Data to Knowledge
- Process of KDD (Knowledge Discovery in Databases)
- Related Technologies
- Comparisons
Why KDD?

We are drowning in information, but starving for knowledge —— John Naisbett

Growing Gap between Data Generation and Data Understanding:

- Automation of business activities:
  - Telephone calls, credit card charges, medical tests, etc.
- Earth observation satellites:
  - Estimated will generate one terabyte ($10^{15}$ bytes) of data per day. At a rate of one picture per second.
- Biology:
  - Human Genome database project has collected over gigabytes of data on the human genetic code [Fasman, Cuticchia, Kingsbury, 1994.]
- US Census data:
- NASA databases:
- …
- World Wide Web:
Process of KDD

Process of KDD

1. Selection
   - Learning the application domain
   - Creating a target dataset

2. Pre-Processing
   - Data cleaning and preprocessing

3. Transformation
   - Data reduction and projection

4. Data Mining
   - Choosing the functions and algorithms of data mining
   - Association rules, classification rules, clustering rules

5. Interpretation and Evaluation
   - Validate and verify discovered patterns

6. Using discovered knowledge
Typical Data Mining Tasks

- **Finding Association Rules** [Rakesh Agrawal et al, 1993]
  - Each transaction is a set of items.
    - Given a set of transactions, an association rule is of the form $X \Rightarrow Y$
    - where $X$ and $Y$ are sets of items.
      - e.g.: 30% of transactions that contain beer also contain diapers;
      - 2% of all transactions contain both of these items.

**Applications:**
- Market basket analysis and cross-marketing
- Catalog design
- Store layout
- Buying patterns
Finding Sequential Patterns

- Each data sequence is a list of transactions.
- Find all sequential patterns with a user-specified minimum support.
  - e.g.: Consider a book-club database
  - A sequential pattern might be
    - 5% of customers bought “Harry Potter I”, then “Harry Potter II”, and then “Harry Potter III”.

Applications:

- Add-on sales
- Customer satisfaction
- Identify symptoms/diseases that precede certain diseases
• **Finding Classification Rules**
  - Finding discriminant rules for objects of different classes.
  - Approaches:
    - Finding Decision Trees
    - Finding Production Rules

**Applications:**
- Process loans and credit cards applications
- Model identification
- Text Mining
- Web Usage Mining
- Etc.
Related Technologies

- **Database Systems**
  - MS SQL server
    - Transaction databases
    - OLAP (Data Cubes)
  - Data Mining
    - Decision Trees
    - Clustering Tools

- **Machine Learning/Data Mining Systems**
  - CART (Classification And Regression Trees)
  - C 5.x (Decision Trees)
  - WEKA (Waikato Environment for Knowledge Analysis)
  - LERS
  - ROSE 2

- **Rule-Based Expert System Development Environments**
  - CLIPS, JESS
  - EXSYS

- **Web-based Platforms**
  - Java
  - MS .Net
## Comparisons

<table>
<thead>
<tr>
<th></th>
<th>Pre-Processing</th>
<th>Learning Data Mining</th>
<th>Inference Engine</th>
<th>End-User Interface</th>
<th>Web-Based Access</th>
<th>Reasoning with Uncertainties</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS SQL Server</td>
<td>N/A</td>
<td>Decision Trees</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>CART C 5.x</td>
<td>N/A</td>
<td>Decision Trees</td>
<td>Built-in</td>
<td>Embedded</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>WEKA</td>
<td>Yes</td>
<td>Trees, Rules,</td>
<td>N/A</td>
<td>Embedded</td>
<td>Need Programming</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clustering,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Association</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLIPS JESS</td>
<td>N/A</td>
<td>N/A</td>
<td>Built-in</td>
<td>Embedded</td>
<td>Need Programming</td>
<td>3rd parties Extensions</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Rough Set Theory

- Introduced by Zdzislaw Pawlak in 1982
- Representative Publications:
- RST is useful for reasoning about knowledge of objects represented by attributes (features).
- Fundamental Assumptions:
  - Objects are represented by values of attributes.
  - Objects with the same information are indiscernible.
Basic Concepts

- **Approximation Space**
  An approximation space is a pair \((U, R)\) where
  \(U\) is a nonempty finite set called the universe and
  \(R\) is an equivalence relation defined on \(U\).

- **Information System**
  An information system is a pair \(S = (U, A)\), where
  \(U\) is a nonempty finite set called the universe and
  \(A\) is a nonempty finite set of attributes, i.e., \(a: U \rightarrow V_a\) for \(a \in A\),
  where \(V_a\) is called the domain of \(a\).

- **Decision Table (Data Table)**
  A decision table is a special case of information systems,
  \(S = (U, A = C \cup \{d\})\),

  where attributes in \(C\) are called **condition attributes** and \(d\) is a designated attribute called the **decision attribute**.
Table 1. Example of Information Table.

<table>
<thead>
<tr>
<th>STUDENT</th>
<th>Category</th>
<th>Major</th>
<th>Birth Place</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PhD</td>
<td>History</td>
<td>Detroit</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>MS</td>
<td>Chemistry</td>
<td>Akron</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>MS</td>
<td>History</td>
<td>Detroit</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>BS</td>
<td>Math</td>
<td>Detroit</td>
<td>B</td>
</tr>
<tr>
<td>5</td>
<td>BS</td>
<td>Chemistry</td>
<td>Akron</td>
<td>C</td>
</tr>
<tr>
<td>6</td>
<td>PhD</td>
<td>Computing</td>
<td>Cleveland</td>
<td>A</td>
</tr>
<tr>
<td>7</td>
<td>BS</td>
<td>Chemistry</td>
<td>Cleveland</td>
<td>C</td>
</tr>
<tr>
<td>8</td>
<td>PhD</td>
<td>Computing</td>
<td>Akron</td>
<td>A</td>
</tr>
</tbody>
</table>
Another Example of Information Table

<table>
<thead>
<tr>
<th>Case</th>
<th>Temperature</th>
<th>Headache</th>
<th>Nausea</th>
<th>Cough</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>high</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>2</td>
<td>very_high</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>3</td>
<td>high</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>4</td>
<td>high</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>5</td>
<td>normal</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>6</td>
<td>normal</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>
## Example of Decision Tables

<table>
<thead>
<tr>
<th>Case</th>
<th>Temperature</th>
<th>Headache</th>
<th>Nausea</th>
<th>Cough</th>
<th>Flu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>high</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>2</td>
<td>very_high</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>3</td>
<td>high</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>4</td>
<td>high</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>5</td>
<td>normal</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>6</td>
<td>normal</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>

### Expanded Table

<table>
<thead>
<tr>
<th>Case</th>
<th>Temperature</th>
<th>Headache</th>
<th>Nausea</th>
<th>Cough</th>
<th>Flu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>high</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>2</td>
<td>very_high</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>3</td>
<td>high</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>4</td>
<td>high</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>5</td>
<td>normal</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>6</td>
<td>normal</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>7</td>
<td>normal</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>8</td>
<td>normal</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>
Approximations of Sets
Let $S = (U, R)$ be an approximation space and $X$ be a subset of $U$.

The lower approximation of $X$ by $R$ in $S$ is defined as

$$\underline{R}X = \{ e \in U \mid [e] \subseteq X \}$$

and

The upper approximation of $X$ by $R$ in $S$ is defined as

$$\overline{R}X = \{ e \in U \mid [e] \cap X \neq \emptyset \}$$

where $[e]$ denotes the equivalence class containing $e$. $[e]$ is called elementary set.
A subset $X$ of $U$ is said to be *R-definable* in $S$ if and only if $R^X = \overline{R^X}$.

The **boundary set** $BN_R(X)$ is defined as $\overline{R^X} - \underline{R^X}$.

A set $X$ is **rough** in $S$ if its boundary set is nonempty.
Approximation Space

Set X

\[ \text{BN}(X) = \overline{RX} - RX \]
Set X
Accuracy of Approximations

$$\alpha_B(X) = \frac{|B(X)|}{|B(X)|}$$
Let \( X = \{(\text{Grade}, A)\} = \{1, 2, 6, 8\} \)

Let \( B = \{\text{Major, Birth\_Place}\} \)

\[
\begin{align*}
\text{U\backslash B} &= \{\{1, 3\}, \{2, 5\}, \{4\}, \{6\}, \{7\}, \{8\}\} \\
B(X) &= \{6, 8\} \\
\bar{B}(X) &= \{1, 2, 3, 5, 6, 8\} \\
\text{BN}_B(X) &= \{1, 2, 3, 5\}
\end{align*}
\]

\[
\alpha_B(X) = \frac{2}{6} = \frac{1}{3}
\]

\[
\begin{array}{|c|c|c|c|c|}
\hline
\text{U} & \text{Category} & \text{Major} & \text{Birth\_Place} & \text{Grade} \\
\hline
1 & \text{PhD} & \text{History} & \text{Detroit} & A \\
2 & \text{MS} & \text{Chemistry} & \text{Akron} & A \\
3 & \text{MS} & \text{History} & \text{Detroit} & C \\
4 & \text{BS} & \text{Math} & \text{Detroit} & B \\
5 & \text{BS} & \text{Chemistry} & \text{Akron} & C \\
6 & \text{PhD} & \text{Computing} & \text{Cleveland} & A \\
7 & \text{BS} & \text{Chemistry} & \text{Cleveland} & C \\
8 & \text{PhD} & \text{Computing} & \text{Akron} & A \\
\hline
\end{array}
\]
Dependency of Attributes

Let $C$ and $D$ be subsets of $A$. We say that $D$ depends on $C$ in a degree $k$ ($0 \leq k \leq 1$) denoted by $C \rightarrow_k D$ if

$$k = \gamma(C, D) = \sum_{X \in U / D} \frac{|C(X)|}{|U|}.$$  

$$k = \gamma(C, D) = \frac{|POS_C(D)|}{|U|}$$

where $POS_C(D) = \bigcup_{X \in U / D} C(X)$, called $C$-positive region of $D$.

If $k = 1$ we say that $D$ depends totally on $C$. 
If $k < 1$ we say that $D$ depends partially (in a degree $k$) on $C$. 

26
<table>
<thead>
<tr>
<th>No.</th>
<th>Category</th>
<th>Major</th>
<th>Birth Place</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PhD</td>
<td>History</td>
<td>Detroit</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>MS</td>
<td>Chemistry</td>
<td>Akron</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>MS</td>
<td>History</td>
<td>Detroit</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>BS</td>
<td>Math</td>
<td>Detroit</td>
<td>B</td>
</tr>
<tr>
<td>5</td>
<td>BS</td>
<td>Chemistry</td>
<td>Akron</td>
<td>C</td>
</tr>
<tr>
<td>6</td>
<td>PhD</td>
<td>Computing</td>
<td>Cleveland</td>
<td>A</td>
</tr>
<tr>
<td>7</td>
<td>BS</td>
<td>Chemistry</td>
<td>Cleveland</td>
<td>C</td>
</tr>
<tr>
<td>8</td>
<td>PhD</td>
<td>Computing</td>
<td>Akron</td>
<td>A</td>
</tr>
</tbody>
</table>

\[ U \setminus \{\text{Category}\} = \{\{1, 6, 8\}, \{2, 3\}, \{4, 5, 7\}\} \]
\[ U \setminus \{\text{Major}\} = \{\{1, 3\}, \{2, 5, 7\}, \{4\}, \{6, 8\}\} \]
\[ U \setminus \{\text{Birth\_Place}\} = \{\{2, 5, 8\}, \{1, 3, 4\}, \{6, 7\}\} \]
\[ U \setminus \{\text{Grade}\} = \{\{1, 2, 6, 8\}, \{4\}, \{3, 5, 7\}\} \]
Dispensable and Indispensable Attributes

Let $S = (U, A = C \cup D)$ be a decision table. Let $c$ be an attribute in $C$. Attribute $c$ is dispensable in $S$ if 

$$POS_c(D) = POS_{(C \setminus \{c\})}(D)$$

Otherwise, $c$ is indispensable.

A decision table $S$ is independent if all attributes in $C$ are indispensable.
Reducts and Core

Let \( S = (U, A = C \cup D) \) be a decision table.
A subset \( R \) of \( C \) is a reduct of \( C \), if

- \( \text{POSR}(D) = \text{POSC}(D) \) and
- \( S' = (U, R \cup D) \) is independent, i.e., all attributes in \( R \) are indispensable in \( S' \).

Core of \( C \) is the set of attributes shared by all reducts of \( C \).

\[
\text{CORE}(C) = \cap \text{RED}(C)
\]

where \( \text{RED}(C) \) is the set of all reducts of \( C \).
<table>
<thead>
<tr>
<th>U</th>
<th>Category</th>
<th>Major</th>
<th>Birth_Place</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PhD</td>
<td>History</td>
<td>Detroit</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>MS</td>
<td>Chemistry</td>
<td>Akron</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>MS</td>
<td>History</td>
<td>Detroit</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>BS</td>
<td>Math</td>
<td>Detroit</td>
<td>B</td>
</tr>
<tr>
<td>5</td>
<td>BS</td>
<td>Chemistry</td>
<td>Akron</td>
<td>C</td>
</tr>
<tr>
<td>6</td>
<td>PhD</td>
<td>Computing</td>
<td>Cleveland</td>
<td>A</td>
</tr>
<tr>
<td>7</td>
<td>BS</td>
<td>Chemistry</td>
<td>Cleveland</td>
<td>C</td>
</tr>
<tr>
<td>8</td>
<td>PhD</td>
<td>Computing</td>
<td>Akron</td>
<td>A</td>
</tr>
</tbody>
</table>

\[
U\{\text{Category}\} = \{1, 6, 8\}, \{2, 3\}, \{4, 5, 7\}
\]

\[
U\{\text{Major}\} = \{1, 3\}, \{2, 5, 7\}, \{4\}, \{6, 8\}
\]

\[
U\{\text{Birth\_Place}\} = \{2, 5, 8\}, \{1, 3, 4\}, \{6, 7\}
\]

\[
U\{\text{Grade}\} = \{1, 2, 6, 8\}, \{4\}, \{3, 5, 7\}
\]

\[
C = \{\text{Category, Major, Birth\_Place}\}
\]

\[
D = \{\text{Grade}\}
\]

\[
U\setminus C = \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\}
\]

\[
C_1 = \{\text{Category, Major}\}, C_2 = \{\text{Category, Birth\_Place}\}
\]

\[
C_3 = \{\text{Major, Birth\_Place}\}
\]

\[
U\setminus C_1 = \{1\}, \{2\}, \{3\}, \{4\}, \{5, 7\}, \{6, 8\}
\]

\[
U\setminus C_2 = \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\}
\]

\[
U\setminus C_3 = \{1\}, \{3\}, \{2, 5\}, \{4\}, \{6\}, \{7\}, \{8\}
\]

\[
\text{POS}_{C}(D) = \{1, 2, 3, 4, 5, 6, 7, 8\}
\]

\[
\text{POS}_{C_1}(D) = \{1, 2, 3, 4, 5, 6, 7, 8\}
\]

\[
\text{POS}_{C_2}(D) = \{1, 2, 3, 4, 5, 6, 7, 8\}
\]

\[
\text{POS}_{C_3}(D) = \{4, 6, 7, 8\}
\]

C1 and C2 are reducts of C.

C3 is not a reduct of C.

The Core of C is \{Category\}.
Rough Membership Function [Pawlak & Skowron, 1994]

Let $S = (U, A)$, $B \subseteq A$ and $X \subseteq U$

Then the rough membership function $\mu^B_X$ for $X$ is a mapping from $U$ to $[0, 1]$, 

$$\mu^B_X : U \to [0, 1]$$

For all $e$ in $U$, the degree of $e$ belongs to $X$ in light of the set of attributes $B$ is defined as 

$$\mu^B_X(e) = |B(e) \cap X| / |B(e)|$$

where $B(e)$ denotes the block containing $e$. 
Properties of rough membership function:

P1: \( \mu^B_X(e) = 1 \) iff \( e \) in \( B_*(X) \)

P2: \( \mu^B_X(e) = 0 \) iff \( e \) in \( U - B^*(X) \)

P3: 0 < \( \mu^B_X(e) < 1 \) iff \( e \) in \( BN_B(X) \)

P4: \( \mu^B_{U \setminus X}(e) = 1 - \mu^B_X(e) \) for any \( e \) in \( U \)

P5: \( \mu^B_{X \cup Y}(e) \geq \max( \mu^B_X(e), \mu^B_Y(e) ) \) for any \( e \) in \( U \)

P6: \( \mu^B_{X \cap Y}(e) \leq \min( \mu^B_X(e), \mu^B_Y(e) ) \) for any \( e \) in \( U \)

where \( B_*(X) \) is the lower approximation of \( X \) in \( B \) and \( B^*(X) \) is the upper approximation of \( X \) in \( B \).
Fuzzy Sets and Rough Sets:

Let \( U \) be a domain of objects.
A fuzzy set \( X \) defined on \( U \) is characterized by a membership function \( \mu_X: \)
\[
\mu_X: U \rightarrow [0, 1]
\]
Let \( A \) and \( B \) be two fuzzy sets, and
\[
\mu_{A \cap B} = \min(\mu_A, \mu_B)
\]
\[
\mu_{A \cup B} = \max(\mu_A, \mu_B)
\]

Let \( S = (U, R) \) be an approximation space and \( X \) be a subset of \( U \).
Define
\[
\mu_X(e) = \begin{cases} 
1 & \text{if } e \text{ in } RX \\
1/2 & \text{if } e \text{ in } \text{BN}_R(X) \\
0 & \text{if } e \text{ in } -\overline{R}X 
\end{cases}
\]
where \(-X\) is the complement of \( X \).

Then, the rough membership function cannot be extended to the fuzzy union and intersection of sets.
In general:

\[
R(X \cup Y) \supseteq R_X \cup R_Y \quad \text{and} \quad \overline{R}(X \cap Y) \subseteq \overline{R}X \cap \overline{R}Y.
\]

The rough membership function will reduce to fuzzy set when

\[
R(X \cup Y) = R_X \cup R_Y \quad \text{and} \quad \overline{R}(X \cap Y) = \overline{R}X \cap \overline{R}Y
\]
Rough MultiSets and MultiSet Decision Tables

Related Concepts:
Rough Sets and Information Systems [Pawlak, 1982]
Rough MultiSets and Information Multisystems [Grzymala-Busse, 1987]
Multiset Decision Tables [Chan 2001, 2004]


<table>
<thead>
<tr>
<th>U</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>E</th>
<th>F</th>
<th>D</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>
The projection of the multirelation onto the set $P$ of attributes $\{A, B, C, E, F\}$ is shown in Table 3.

Table 3. An information multisystem

<table>
<thead>
<tr>
<th>$A$</th>
<th>$B$</th>
<th>$C$</th>
<th>$E$</th>
<th>$F$</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4. A sub-multiset $X$ of $\tilde{P}$.

<table>
<thead>
<tr>
<th>$A$</th>
<th>$B$</th>
<th>$C$</th>
<th>$E$</th>
<th>$F$</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5. $P$-lower approximation of $X$.

<table>
<thead>
<tr>
<th>$A$</th>
<th>$B$</th>
<th>$C$</th>
<th>$E$</th>
<th>$F$</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 6. $P$-upper approximation of $X$.

<table>
<thead>
<tr>
<th>$A$</th>
<th>$B$</th>
<th>$C$</th>
<th>$E$</th>
<th>$F$</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>
The classification of $P$ generated by attribute $D$ in $S$:

consists of three sub-multisets which are given Tables 7, 8, and 9.

Table 7. Sub-multiset of the multipartition $D_P$ with $D = 1$.

<table>
<thead>
<tr>
<th>$A$</th>
<th>$B$</th>
<th>$C$</th>
<th>$E$</th>
<th>$F$</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 8. Sub-multiset of the multipartition $D_P$ with $D = 2$.

<table>
<thead>
<tr>
<th>$A$</th>
<th>$B$</th>
<th>$C$</th>
<th>$E$</th>
<th>$F$</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 9. Sub-multiset of the multipartition $D_P$ with $D = 3$.

<table>
<thead>
<tr>
<th>$A$</th>
<th>$B$</th>
<th>$C$</th>
<th>$E$</th>
<th>$F$</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Multiset Decision Table

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>E</th>
<th>F</th>
<th>W</th>
<th>D₁</th>
<th>D₂</th>
<th>D₃</th>
<th>w₁</th>
<th>w₂</th>
<th>w₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Condition attributes $C = \{A, B, C, E, F\}$ and decision attribute $D$.

The Boolean vector is denoted by $[D1, D2, D3]$, and the integer vector is denoted by $[w1, w2, w3]$. Note that $W = w1 + w2 + w3$ on each row.
To determine the partition of boundary multisets, we use the following two steps.

**Step 1.** Identify rows with $D1 + D2 + D3 > 1$, we have the following multiset in table form:

**Table 11. C-lower approximation of $D1$.**

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>E</th>
<th>F</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 12. C-upper approximation of $D1$.**

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>E</th>
<th>F</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

**Table 13. Elements in the boundary sets.**

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>E</th>
<th>F</th>
<th>W</th>
<th>$D_1$</th>
<th>$D_2$</th>
<th>$D_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
To determine the partition of boundary multisets, we use the following two steps.

**Step 1.** Identify rows with $D1 + D2 + D3 > 1$, we have the following multiset in table form:

**Table 13.** Elements in the boundary sets.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>E</th>
<th>F</th>
<th>W</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

**Step 2.** Grouping the above table in terms of $D1$, $D2$, and $D3$, we have the following blocks in the partition.
The relationship between rough set theory and Dempster-Shafer’s theory of evidence was first shown in [Grzymala-Busse, 1987] and further developed in [Skowron, A. and J. Grzymala-Busse, 1994].

The concept of partition of boundary sets was introduced in [Skowron, A. and J. Grzymala-Busse, 1994].

The basic idea is to represent an expert’s classification on a set of objects in terms of lower approximations and a partition on the boundary set.


Table 18. Grouping over $D_1, D_2, D_3$ and sum over $W$.

<table>
<thead>
<tr>
<th>$D_1$</th>
<th>$D_2$</th>
<th>$D_3$</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

Let $\Theta = \{1, 2, 3\}$.

Table 19. The bpa derived from Table 2.

<table>
<thead>
<tr>
<th>X</th>
<th>${1}$</th>
<th>${2}$</th>
<th>${3}$</th>
<th>${1, 2}$</th>
<th>${1, 3}$</th>
<th>${2, 3}$</th>
<th>${1, 2, 3}$</th>
</tr>
</thead>
</table>
Rough Sets and Bayes’ Theorem
[Pawlak 1999, 2002]
Rough Membership Function [Pawlak & Skowron, 1994]

Let $S = (U, A)$, $B \subseteq A$ and $X \subseteq U$
Then the rough membership function $\mu_B X$ for $X$ is a mapping
from $U$ to $[0, 1]$,

$$\mu_B X : U \rightarrow [0, 1]$$

For all $e$ in $U$, the degree of $e$ belongs to $X$ in light of the set of
attributes $B$ is defined as

$$\mu_B X(e) = \frac{|B(e) \cap X|}{|B(e)|}$$

where $B(e)$ denotes the block containing $e$. 
Let $C \rightarrow D$ be a decision rule and let $e$ in $U$, then

**Support** of $C \rightarrow_e D$ is defined as
\[
\text{supp}_e(C, D) = |C(e) \cap D(e)|
\]

**Strength** of $C \rightarrow_e D$:
\[
\sigma_e(C, D) = \frac{|C(e) \cap D(e)|}{|U|} = \frac{\text{supp}_e(C, D)}{|U|}
\]

**Certainty** of $C \rightarrow_e D$:
\[
\text{cer}_e(C, D) = \frac{|C(e) \cap D(e)|}{|C(e)|}
\]

**Inverse decision rule:**
Let $C \rightarrow_e D$ be a decision rule, then $D \rightarrow_e C$ is an inverse decision rule of $C \rightarrow_e D$.

**Coverage** of $C \rightarrow_e D$: (certainty of the $D \rightarrow_e C$)
\[
\text{cove}(C, D) = \frac{|D(e) \cap C(e)|}{|D(e)|}
\]
Learning Rules from Examples

- **LEM2** [Chan, 1989, 1991]
  - Basic idea is to learn rules from lower and upper approximations
  - Work in incremental and non-incremental modes

- **RLEM2** [Chan, 2001]
  - Basic idea is to learn rules from Multiset Decision Tables using extended SQL operators
  - Rules are learned from lower and upper approximations

- **BLEM2** [Chan, 2003]
  - Basic idea is to learn rules from lower approximation and partition of boundary set
  - Rules are associated with four factors based on Bayes’ theorem and rough sets [Pawlak, 2002]
Generate rules with support, certainty, strength, and coverage

For a decision rule \( r, T \rightarrow (d, v) \), derived from a decision table \((U, A=C \cup D)\), we have

- **support** of \( r \), \( \text{supp}(r) = |[T] \cap [(d, v)]| \),
- **strength** of \( r \), \( \sigma(r) = |[T] \cap [(d, v)]| / |U| \),
- **certainty** of \( r \), \( \text{cer}(r) = |[T] \cap [(d, v)]| / |[T]| \), and
- **coverage** of \( r \), \( \text{cov}(r) = |[T] \cap [(d, v)]| / |[(d,v)]| \).

where \( U \) is a finite set of examples,
\( C \) is a set of condition attributes and
\( D \) is a singleton set \( \{d\} \) of decision attribute,
\((d, v)\) is a decision-value pair,
\( T \) is a nonempty set of condition-value pairs,
\([T] = \bigcap_{t \in T} [t] \) is the block of \( T \).
$T \rightarrow (d, v)$ if and only if $[T] \subseteq [(d, v)]$

Set $T$ is called a **complex** of $(d, v)$ when $T \rightarrow (d, v)$

Set $T$ is a **minimal complex** of $(d, v)$ when $T$ is minimal, i.e., for any $t$ in $T$, $[T - \{t\}] \not\subset [(d, v)]$

A nonempty set $R$ of minimal complexes of $(d, v)$ is a **local covering** of $(d, v)$, if $\bigcup_{T \in R} [T] = [(d, v)]$ and $R$ is minimal.

The objective is to find a local covering for each decision-value pair in a decision table.
procedure LEM2
inputs: set X, which is a lower or upper approximation of a decision-value pair.
outputs: a single local covering T of the decision-value pair.
begin
  G := X; //initial target set
  T := ∅; //final local covering
  while G ≠ ∅ do
    begin
      T := ∅; //initial complex
      T(G) := {t | [t] ∩ G ≠ ∅ }; //list of relevant a-v pairs
      while T = ∅ or not ([T] ⊆ X) do
        begin
          select a pair t in T(G) with the highest rank, if a tie occurs, select a pair t in T(G) such that |[t] ∩ G| is maximum; if another tie occurs, select a pair t in T(G) with the smallest cardinality of [t]; if a further tie occurs, select the first pair;
          T := T ∪ {t};
          G := [t] ∩ G;
          T(G) := {t | [t] ∩ G ≠ ∅};
          T(G) := T(G) − T;
        end; //while
      for each t in T do
        if [T − {t}] ⊆ X then T := T − {t}; //LINE 21
        T := T ∪ {T};
        G := X − ∪ T ∈ T[T];
      end; //while
    for each T ∈ T do
      if ∪ S ∈ T − {T}[S] = X then T := T − {T};
    end; {procedure LEM2}
BLEM2:
if \([T - \{t\}] \subseteq G\) then \(T := T - \{t\};\)  

//LINE 21

**LEM2 generates smaller number of rules than BLEM2.**

BLEM2 generates rules with support, strength, certainty, and coverage factors.

Facilitate the design of Bayesian classifiers using BLEM2 rules.

Experimental results showed that they are effective tools for developing efficient inference engines.
Tools Based on Rough Set

*ROSE/ROSE*₂
ROugh Set data Explorer

*4eMka*
Dominance-based Rough Set Approach to Multicriteria Classification

*JAMM*
A New Decision Support Tool for Analysis and Solving of Multicriteria Classification Problems

http://idss.cs.put.poznan.pl/site/software.html
LERS

• A data mining system LERS (Learning from Examples based on Rough Sets) computes lower and upper approximations of all concepts,

• Rules induced from the lower approximation of the concept are called certain,

• Rules induced from the upper approximation of the concept are called possible,

• LEM2 explores the search space of attribute-value pairs,

• MLEM2, a modified version of LEM2.
Some Applications of LERS

• Used by NASA Johnson Space Center (Automation and Robotics Division), a tool to develop expert systems for medical decision-making on board of the Space Station Freedom

• Enhancing facility compliance under Sections 311, 312, and 313 of Title III, the Emergency Planning and Community Right to Know

• Assessing preterm labor risk for pregnant women
  • Currently used manual methods of assessing preterm birth have a positive predictive value of 17–38%.
  • The data mining methods based on LERS reached positive predictive value of 59–92%,
• diagnosis of melanoma,
• prediction of behavior under mental retardation,
• analysis of animal models for prediction of self-injurious behavior,
• nursing,
• global warming,
• natural language, and
• data transmission.
# Performance of AQ15, C4.5 and LERS Error Rates

<table>
<thead>
<tr>
<th>Data Set</th>
<th>AQ15</th>
<th>C4.5</th>
<th>LERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lymphography</td>
<td>18–20%</td>
<td>23%</td>
<td>19%</td>
</tr>
<tr>
<td>Breast Cancer</td>
<td>32–34%</td>
<td>33%</td>
<td>30%</td>
</tr>
<tr>
<td>Primary Tumor</td>
<td>59–71%</td>
<td>60%</td>
<td>67%</td>
</tr>
</tbody>
</table>
Rule-Based Data Mining System Objectives

- Develop an integrated rule-based data mining system provides
  - Synergy of database systems, machine learning, and expert systems
  - Dealing with uncertain rules
  - Delivery of web-based user interface
Structure of Rule-Based Systems
Proposed System Workflow

Input Data Set → Data Pre-processing → Rule Generator → User Interface Generator
Input Data Set:
- Text file with comma separated values (CSV)
- It is assumed that there are N columns of values corresponding to N variables or parameters, which may be real or symbolic values.
- The first N – 1 variables are considered as inputs and the last one is the output variable.

Data Preprocessing:
- Discretize domains of real variables into a finite number of intervals
- Discretized data file is then used to generate an attribute information file and a training data file.

Rule Generator:
- A symbolic learning program called BLEM2 is used to generate rules with uncertainty

User Interface Generator:
- Generate a web-based rule-based system from a rule file and corresponding attribute file
Architecture of RBC generator

Workflow of RBC generator

- Rule set File
- Metadata File
- RBC Generator
- SQL Rule Table
- Rule Table Definition
Benefits

A system for generating rule-based classifier from data with the following benefits:

- No need of end user programming
- Automatic rule-based system creation
- Delivery system is web-based provides easy access
Rough Set Exploration System

What is RSES?
RSES is a toolkit for analysis of table data running under Windows NT/95/98/2000/XP. It is based on methods and algorithms coming from the area of Rough Sets. It comprises of two general components - the GUI front-end and the computational kernel. The kernel is based on renewed RSESlib library.

Requirements:
PC with 128+ MB RAM.
3 MB of disc space + space occupied by Java VM
Windows NT4/95/98/2000/XP or Linux/i386
Java Runtime Environment (JRE) or Java SDK. We recommend the use of version 1.4.1 or higher.

The system, starting from version 2.0, is distributed as single self-installing bundle (for Windows). The installation procedure is described in User's Guide. [http://logic.mimuw.edu.pl/~rses](http://logic.mimuw.edu.pl/~rses)

Computational algorithms implemented by (in alphabetical order): Jan Bazan, Rafał Latkowski, Nguyen Sinh Hoa, Nguyen Hung Son, Piotr Synak, Arkadiusz Wojna, Marcin Wojnarski and Jakub Wróblewski.

Copyright © 1994-2005 Logic Group, Institute of Mathematics, Warsaw University, Poland.
Concluding Remarks

- Classical Rough Set Theory (CRST) introduced by Pawlak has been applied to the development of learning and data reduction algorithms for data mining tasks.
- Extensions of CRST such as Dominance Based Rough Sets (DBRS) further facilitate the development of tools for Multi-Criteria Decision Analysis.
- Rough Sets + Genetic Algorithms still have rooms to be developed.
- Extension from rough sets to granular computing is undergoing development.
Thank You!
References

