A granular computing approach to machine learning

2010.09.30

Eung-Hee Kim
Agenda

• Abstract
• Introduction
• A granular computing view of data mining
• A typical machine learning problem
• Classification algorithms
• Conclusion
Abstract

• **Goal of Granular Computing**
  – Using of granules in problem solving

• **Classification**
  – intensely studied problems in machine learning

• **Aim of this paper**
  – Proposing a granular computing approach for classification
    • By studying & extending ID3 and PRISM algorithm
Introduction

• Knowledge discovery & Data mining
  – Extraction of interesting info & patterns from large databases
  – Major goal
    • Learning & Identification for knowledge, patterns and regularities

• Representation of knowledge
  – Rules / Black box systems (in ANN)

• Lots of studies have focused on
  – Algorithms for many different rules
  – Speeding up of existing algorithms
Introduction

• In logicians’ view point

Data mining (esp., rule mining)

Formation of concepts

Identification of relationship between concepts

Concept (granule)

Intension

A set of properties (attributes)

Extension (granule)

A set of entities (objects)
Introduction

Granular Computing

- methods
- theories
- tools
- techniques

Using Granules
(Subset of universe)

- subsets
- classes
- clusters

Concept := (extent, intent) = Unit of thought

An example of granular computing data mining model

- Result from FCA
- Result from Granular computing

- Concept
- Extent
- Intent
- Granule
- Description
Introduction

Granulation cases

Partition

Covering

ID3

PRISM
A granular computing view of data mining

- Granule: a subset of universe
  - \{A\}, \{A, B\}, \{A, B, C\}, ...

- Granulation: a family of granules containing every objects in the universe
  - \{\{A, B, C\}, \{D, E\}\}, \{\{A, B\}, \{B, C, D\}, \{E\}\}, ...

- Partition, Covering \(\subseteq\) Granulation
  - Partition:= disjoint subsets of the universe
    - \{\{A, B, C\}, \{D, E\}\}, ...
  - Covering:= subsets of the universe allowing overlapping
    - \{\{A, B\}, \{B, C, D\}, \{E\}\}
    - Partition \(\subseteq\) Covering
  - \(\therefore\) Granule = each element of partitions or coverings
Basic definitions & symbols

• Information table
  – \( S=(U, A_t, L, \{V_a| a \in A_t\}, \{I_a| a \in A_t\}) \)
  • \( U \): a finite nonempty set of objects
    – \( \{o_1, o_2, o_3, \ldots, o_8\} \)
  • \( A_t \): a finite nonempty set of attributes
    – \( \{\text{height, hair, eyes}\} \)
  • \( L \): a language defined using attributes in \( A_t \)
  • \( V_a \): a nonempty set of values for \( a \in A_t \)
    – \( V_{\text{hair}}=\{\text{blond, red, dark}\} \)
  • \( I_a \): an information function \( U \rightarrow V_a \)
  • \( I_A(x) \) where \( A \subseteq A_t \) and \( x \in U \): the value of an object \( x \) on \( A \)
    – \( I_{\{\text{height}\}}(o_1)=[\text{short}] \), \( I_{\{\text{height, hair, eyes}\}}(o_2)=[\text{short, blond, brown}] \)

<table>
<thead>
<tr>
<th>object</th>
<th>height</th>
<th>hair</th>
<th>eyes</th>
</tr>
</thead>
<tbody>
<tr>
<td>o₁</td>
<td>short</td>
<td>blond</td>
<td>blue</td>
</tr>
<tr>
<td>o₂</td>
<td>short</td>
<td>blond</td>
<td>brown</td>
</tr>
<tr>
<td>o₃</td>
<td>tall</td>
<td>red</td>
<td>blue</td>
</tr>
<tr>
<td>o₄</td>
<td>tall</td>
<td>dark</td>
<td>blue</td>
</tr>
<tr>
<td>o₅</td>
<td>tall</td>
<td>dark</td>
<td>blue</td>
</tr>
<tr>
<td>o₆</td>
<td>tall</td>
<td>blond</td>
<td>blue</td>
</tr>
<tr>
<td>o₇</td>
<td>tall</td>
<td>dark</td>
<td>brown</td>
</tr>
<tr>
<td>o₈</td>
<td>short</td>
<td>blond</td>
<td>brown</td>
</tr>
</tbody>
</table>
Language L

• Language L
  – Atomic formula: \( a = v \) where \( a \in A_t \) and \( v \in V_a \)
    ♦ hair = blond, eyes=blue
  – Logical operators: negation, conjunction and disjunction
    ♦ Extended formula
      – \( \neg (\text{hair} = \text{blond}) \land (\text{eyes} = \text{blue}) \lor (\text{height} = \text{tall}) \)

• Semantics of the language L
  – The satisfiability of a formula \( \varnothing \) by an object \( x \), written \( x \models S \varnothing \) (in short \( x \models \varnothing \))
    1) \( x \models a = v \) iff \( I_a(x) = v \)
    2) \( x \models \neg \varnothing \) iff not \( x \models \varnothing \)
    3) \( x \models \varnothing \land \psi \) iff \( x \models \varnothing \) and \( x \models \psi \)
    4) \( x \models \varnothing \lor \psi \) iff \( x \models \varnothing \) or \( x \models \psi \)
Concept := (intension, extension)

• If $\emptyset$ is a formula
  – $m_S(\emptyset) = \{ x \in U \mid x \models \emptyset \}$, in short, $m(\emptyset)$
    • $m(\emptyset) \subseteq U$ is the meaning of the formula $\emptyset$
    • $\emptyset$ is the description of $m(\emptyset)$
    • $\therefore$ a connection between formulas of L and subsets of U is established

• Formal description of concepts with the language L
  – A concept :=($\emptyset$, m($\emptyset$)), where $\emptyset \in L$
    • $\emptyset$: intension, description
    • m($\emptyset$): extension, granule

• Granules with the language L
  – For an atomic formula $a = v$
    • m($a = v$)
  – For two formulas $\emptyset$, $\psi \in L$
    • $m(\emptyset) \cap m(\psi) = m(\emptyset \land \psi)$
    • $m(\emptyset) \cup m(\psi) = m(\emptyset \lor \psi)$
Conjunctively definable granule

• Definable granule
  – A set of objects $X \subseteq U$ is definable granule in an information table $S$
    • If there exists a formula $\varnothing \in L$ such that $m(\varnothing) = X$
    – E.g. $\{o_1, o_2, o_3, o_8\}$ is a definable granule
      » $\varnothing$: Height=short \lor hair=red, $m(\varnothing) = \{o_1, o_2, o_3, o_8\}$

• Conjunctively definable granule
  – A set of objects $X \subseteq U$ is conjunctively definable granule in an information table $S$
    • If there exists a formula $\varnothing \in L$ such that $\varnothing$ is a conjunction of atomic formulas and $m(\varnothing) = X$
    – E.g. $\{o_1, o_2, o_8\}$ is a conjunctively definable granule
      » $\varnothing$: height=short $\land$ hair=blond, $m(\varnothing) = \{o_1, o_2, o_8\}$

<table>
<thead>
<tr>
<th>object</th>
<th>height</th>
<th>hair</th>
<th>eyes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$o_1$</td>
<td>short</td>
<td>blond</td>
<td>blue</td>
</tr>
<tr>
<td>$o_2$</td>
<td>short</td>
<td>blond</td>
<td>brown</td>
</tr>
<tr>
<td>$o_3$</td>
<td>tall</td>
<td>red</td>
<td>blue</td>
</tr>
<tr>
<td>$o_4$</td>
<td>tall</td>
<td>dark</td>
<td>blue</td>
</tr>
<tr>
<td>$o_5$</td>
<td>tall</td>
<td>dark</td>
<td>blue</td>
</tr>
<tr>
<td>$o_6$</td>
<td>tall</td>
<td>blond</td>
<td>blue</td>
</tr>
<tr>
<td>$o_7$</td>
<td>tall</td>
<td>dark</td>
<td>brown</td>
</tr>
<tr>
<td>$o_8$</td>
<td>short</td>
<td>blond</td>
<td>brown</td>
</tr>
</tbody>
</table>
Conjunctively definable partition & covering

• A partition $\pi$ is called a conjunctively definable partition
  – If every equivalence class of $\pi$ is a conjunctively definable granule

• A covering $\tau$ is called a conjunctively definable covering
  – If every equivalence class of $\tau$ is a conjunctively definable granule

• To understand the meaning of equivalence class
  – Understanding of equivalence relation should be involved
Equivalence relation

• An equivalence relation is any relation, written $a \sim b$, satisfying three rules
  1. $a \sim a$ (reflexivity)
  2. If $a \sim b$, then $b \sim a$ (symmetry)
  3. If $a \sim b$ and $b \sim c$, then $a \sim c$ (transitivity)

• Example 1. Equals (=) is an equivalence relation
  1. $a = a$ all the time
  2. If $a=b$, $b=a$
  3. If $a=b$ and $b=c$, $a=c$

• Example 2. Let us define $a \sim b$ if $a$ and $b$ have the same sign (0: positive)
  – $a$ has the same sign as $a$, so $a \sim a$
  – If $a$ has the same sign as $b$, $b$ has the same sign as $a$
  – If $a$ has the same sign as $b$, and $b$ has the same sign as $c$, $a$ and $c$ must have the same sign
    • $3 \sim 5, 7 \sim 2, -5 \sim -18$, not $-6 \sim 3$
Equivalence relation & Equivalence class

• Example 3. We will say $a \sim b$ if $a$ and $b$ have the same remainder when you divide by three
  - $3 \sim 6$, $1 \sim 16$, not $2 \sim 7$
  - There are three equivalence classes
    • The things with remainder 0 when you divided by three
      - Equivalence class $1=\{0, 3, 6, 9, \ldots\}$
    • The thins with remainder 1 when you divided by three
      - Equivalence class $2=\{1, 4, 7, 10, \ldots\}$
    • The thins with remainder 2 when you divided by three
      - Equivalence class $3=\{2, 5, 8, 11, \ldots\}$

• Distinguish size of an equivalence class and number of equivalence classes
  - In the case of relation with equals (=), size of an equivalence class = 1, # of equivalence classes = infinite
  - In the case of relation with sign, size of an equivalence class = infinite, # of equivalence classes = 2
  - In the case of relation with mod 3, size of an equivalence class = infinite, # of equivalence classes = 3
Formal definition of Equivalence class
(in WolframMathWorld http://mathworld.wolfram.com)

• An equivalence class is defined as a subset of the form \( \{ x \in X : x \mathrel{R_a} \} \), where \( a \) is an element of \( X \) and the notation “\( x \mathrel{R_y} \)” is used to mean that there is an equivalence relation between \( x \) and \( y \).

• It can be shown that any two equivalence classes are either equal or disjoint, hence the collection of equivalence classes forms a partition of \( X \).

• For all \( a, b \in X \), we have \( a \mathrel{R_b} \) iff \( a \) and \( b \) belong to the same equivalence class.

• A set of class representatives in a subset of \( X \) which contains exactly one element from each equivalence class.
  
  – E.g. For \( n \) a positive integer, and \( a, b \) integers, consider the \( a \equiv b \) (mod \( n \)), then the equivalence classes are the sets \( \{ \ldots, -2n, -n, 0, n, 2n, \ldots \} \), \( \{ \ldots, 1-2n, 1-n, 1, 1+n, 1+2n, \ldots \} \) etc. The standard class representatives are taken to be \( 0, 1, 2, \ldots, n-1 \).
Classification: A typical machine learning problem

- S=(U, A_t, L, \{V_a| a \in A_t\}, \{I_a| a \in A_t\})
  - A_t = F \cup \{\text{class}\}
    - where F is a set of attributes used to describe objects

- The goal of Classification
  - Finding classification rules of the form, \( \emptyset \Rightarrow \text{class} = c_i \), where \( \emptyset \) is a formula over F and \( c_i \) is a class label

- Let \( \pi_{\text{class}} \subseteq \Pi(U) \) denote the partition induced by the attribute class
  - E.g. \( \pi_{\text{class}} = \{\{o_1, o_3, o_6\}, \{o_2, o_4, o_5, o_7, o_8\}\} \)

- An information table provides a consistent classification
  - If \( I_F(x) = I_F(y) \), then \( I_{\text{class}}(x) = I_{\text{class}}(y) \)
    - \( I_F(o_4) = \{\text{tall, dark, blue}\}, I_F(o_5) = \{\text{tall, dark, blue}\} \Rightarrow I_{\text{class}}(o_4) = -, I_{\text{class}}(o_5) = - \)

- Symbol \( \preceq \) (set partial order)
  - \( A = \{\{1, 2, 3\}, \{4, 5, 6, 7\}\}, B = \{\{1, 2\}, \{3\}, \{4, 5\}, \{6, 7\}\} \)
  - \( B \preceq A \)
Consistently classifiable information table

• Formal definition of consistently classifiable information table
  – An information table is a consistent classification problem if and only if
    \[ \pi_F \leq \pi_{\text{class}} \]
    • \( \pi_{\text{class}} = \{ \{ o_1, o_3, o_6 \}, \{ o_2, o_4, o_5, o_7, o_8 \} \} \)
    • \( \pi_F = \{ \{ o_1 \}, \{ o_2, o_8 \}, \{ o_3 \}, \{ o_4, o_5 \}, \{ o_6 \}, \{ o_7 \} \} \)
      – \( \pi_F \leq \pi_{\text{class}} \)

• However, \( \pi_F \) is not very interesting!
  – A subset of attributes from \( F \) producing the correct classification is more interest!
    • A conjunctively definable partition \( \pi \) such that \( \pi \leq \pi_{\text{class}} \)
    • A conjunctively definable covering \( \tau \) such that \( \tau \leq \pi_{\text{class}} \)

• \( X \): a granule in a partition or covering of the universe
• \( \text{des}(X) \): description of \( X \) using language \( L \)
  – If \( X \subseteq m(\text{class} = c_i) \), we can construct a classification rule:
    • \( \text{des}(X) \Rightarrow \text{class} = c_i \)

<table>
<thead>
<tr>
<th>object</th>
<th>height</th>
<th>hair</th>
<th>eyes</th>
<th>class</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>
Partition and Covering

• **Partition**
  – An object is only classified by one rule

• **Covering**
  – An object many be classified by more than one rule

• **Example**
  – \( \pi_{\text{class}} = \{\{o_1, o_3, o_6\}, \{o_2, o_4, o_5, o_7, o_8\}\} \)
  – \( \pi = \{\{o_1, o_6\}, \{o_2, o_8\}, \{o_3\}, \{o_4, o_5, o_7\}\} \): partition
  – \( \tau = \{\{o_1, o_6\}, \{o_2, o_7, o_8\}, \{o_3\}, \{o_4, o_5, o_7\}\} \): covering

  • \( \pi \preceq \pi_{\text{class}} \) and \( \tau \preceq \pi_{\text{class}} \)

<table>
<thead>
<tr>
<th>object</th>
<th>height</th>
<th>hair</th>
<th>eyes</th>
<th>class</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>
Partition

- Example
  
  - $\pi_{\text{class}} = \{\{o_1, o_3, o_6\}, \{o_2, o_4, o_5, o_7, o_8\}\}$
  
  - $\pi = \{\{o_1, o_6\}, \{o_2, o_8\}, \{o_3\}, \{o_4, o_5, o_7\}\}$: partition
  
  - $\tau = \{\{o_1, o_6\}, \{o_2, o_7, o_8\}, \{o_3\}, \{o_4, o_5, o_7\}\}$: covering
    
    - $\pi \preceq \pi_{\text{class}}$ and $\tau \preceq \pi_{\text{class}}$

- Classification rules of $\pi$
  
  - hair=blond $\land$ eyes=blue $\Rightarrow$ class=+: $\{o_1, o_6\}$
  
  - hair=blond $\land$ eyes=brown $\Rightarrow$ class =-: $\{o_2, o_8\}$
  
  - hair=red $\Rightarrow$ class=+: $\{o_3\}$
  
  - hair=dark $\Rightarrow$ class=-: $\{o_4, o_5, o_7\}$

<table>
<thead>
<tr>
<th>object</th>
<th>height</th>
<th>hair</th>
<th>eyes</th>
<th>class</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>
Covering

• Example
  
  $\pi_{\text{class}} = \\{\{o_1, o_3, o_6\}, \{o_2, o_4, o_5, o_7, o_8\}\}$

  $\pi = \{\{o_1, o_6\}, \{o_2, o_8\}, \{o_3\}, \{o_4, o_5, o_7\}\}$: partition

  $\tau = \{\{o_1, o_6\}, \{o_2, o_7, o_8\}, \{o_3\}, \{o_4, o_5, o_7\}\}$: covering
  
  $\bullet \pi \preceq \pi_{\text{class}}$ and $\tau \preceq \pi_{\text{class}}$

  $\pi$ Classification rules of $\pi$

  $\bullet$ hair=red $\Rightarrow$ class=+: $\{o_3\}$

  $\bullet$ eyes=blue $\land$ hair=blond $\Rightarrow$ class=+: $\{o_1, o_6\}$

  $\bullet$ eyes=brown $\Rightarrow$ class =-: $\{o_2, o_7, o_8\}$

  $\bullet$ hair=dark $\Rightarrow$ class=-: $\{o_4, o_5, o_7\}$

<table>
<thead>
<tr>
<th>object</th>
<th>height</th>
<th>hair</th>
<th>eyes</th>
<th>class</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>
Classification algorithms

Study and remodel some popular classification

ID3 for partitioning
PRISM for covering
ID3 (for partition)

• Characteristic of ID3
  – Using of **information gain** to select a suitable attribute to partition the universe
  – Keep previous step until all granules can be understood or expressed by a formula

• What is information gain?!
  – Entropy before information gain
  – Information gain with Building decision tree based ID3 algorithm
Calculation of information before Entropy

• Based on Information Theory
  – The lower the possibility of an event is, the more the event has information
  – Let X is a variable
    • If e is an value of X and P(e) is the possibility \(\in [0, 1]\) of e, the self-information of e is
      - \(h(e) = -\log_2 P(e)\)
        » if \(P(e_1)=1/1024\), \(h(e_1)=-\log_2 2^{-10}=10\) (bit)
        » if \(P(e_2)=1/32\), \(h(e_2)=-\log_2 2^{-5}=5\) (bit)
        » if \(P(e_3)=1\), \(h(e_3)=-\log_2 2^0= 0\) (bit)

• Meaning of \(-P(e)\log_2 P(e)\)
  – Average amount of the information from the value e
Entropy

- For a variable X, the amount of information which can be obtained from the variable X is

\[ H(X) = -\sum_{i=1}^{n} P(e_i) \log_2 P(e_i) \]

and it is called ‘Entropy of the variable X.’

- The case that variable X has the minimum entropy is where
  - There is a value \( e_i \) where \( P(e_i) = 1 \)

- The case that variable X has the maximum entropy is where
  - All the values of the variable X has the same probability
## ID3 algorithm to analyze Simpsons

<table>
<thead>
<tr>
<th>Person</th>
<th>Hair Length</th>
<th>Weight</th>
<th>Age</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homer</td>
<td>0”</td>
<td>250</td>
<td>36</td>
<td>M</td>
</tr>
<tr>
<td>Marge</td>
<td>10”</td>
<td>150</td>
<td>34</td>
<td>F</td>
</tr>
<tr>
<td>Bart</td>
<td>2”</td>
<td>90</td>
<td>10</td>
<td>M</td>
</tr>
<tr>
<td>Lisa</td>
<td>6”</td>
<td>78</td>
<td>8</td>
<td>F</td>
</tr>
<tr>
<td>Maggie</td>
<td>4”</td>
<td>20</td>
<td>1</td>
<td>F</td>
</tr>
<tr>
<td>Abe</td>
<td>1”</td>
<td>170</td>
<td>70</td>
<td>M</td>
</tr>
<tr>
<td>Selma</td>
<td>8”</td>
<td>160</td>
<td>41</td>
<td>F</td>
</tr>
<tr>
<td>Otto</td>
<td>10”</td>
<td>180</td>
<td>38</td>
<td>M</td>
</tr>
<tr>
<td>Krusty</td>
<td>6”</td>
<td>200</td>
<td>45</td>
<td>M</td>
</tr>
</tbody>
</table>
The entropy of the information table

- For the variable X about Sex
  - \( P(e_1) = \frac{5}{9} \) where the value \( e_1 \) is ‘MAN’
  - \( P(e_2) = \frac{4}{9} \) where the value \( e_2 \) is ‘WOMAN’
  - \( H(X) = -P(e_1) \log_2 P(e_1) - P(e_2) \log_2 P(e_2) \)
    \[ = 0.9911 \]

- Our Goal
  - Reducing the Entropy by selecting the most suitable attributes among ‘Hair length, Weight, Age’
Start with ‘Hair length ≤ 5’

Initial status (4F:5M), Entropy (4F:5M) = 0.9911

$\text{Gain}(A) = E(\text{Current set}) - \sum E(\text{all child sets})$

$\text{Gain}(\text{Hair Length } \leq 5) = 0.9911 - (4/9 \times 0.8113 + 5/9 \times 0.9710) = 0.0911$
With ‘Weight ≤ 160’

Initial status (4F:5M), Entropy (4F:5M)=0.9911

Child_1 (4F:1M)

Child_2 (0F:4M)

\[
\text{Entropy}(4F,1M) = -(4/5)\log_2(4/5) - (1/5)\log_2(1/5) \\
= 0.7219
\]

\[
\text{Entropy}(0F,4M) = -(0/4)\log_2(0/4) - (4/4)\log_2(4/4) \\
= 0
\]

\[
\text{Gain}(A) = E(\text{Current set}) - \sum E(\text{all child sets})
\]

\[
\text{Gain}(\text{Weight} \leq 160) = 0.9911 - (5/9 \times 0.7219 + 4/9 \times 0) = 0.5900
\]
With ‘Age ≤40’

Initial status (4F:5M), Entropy (4F:5M)=0.9911

熵计算如下:

熵(4F,2M) = -(3/6)log₂(3/6) - (3/6)log₂(3/6)
= 1

熵(1F,2M) = -(1/3)log₂(1/3) - (2/3)log₂(2/3)
= 0.9183

Gain(A) = E(Current set) - \sum E(all child sets)

Gain(Age <= 40) = 0.9911 - (6/9 * 1 + 3/9 * 0.9183) = 0.0183
Now, select !!

• Information gain comparison
  – Hair ≤5 : 0.0911
  – Weight ≤160: 0.5900
  – Age ≤40 : 0.0183

• The value ‘Weight ≤160’ contributes best to reduce the entropy (⇔ the best information gain) so that it is selected!
Completion of building Decision Tree

• Stop when the entropy of every leaf became ‘zero’
Entropy in Concept Lattice

({p_1, p_2, p_3, p_4, p_5, p_6, p_7}, ø)
Back to the paper

<table>
<thead>
<tr>
<th>object</th>
<th>height</th>
<th>hair</th>
<th>eyes</th>
<th>class</th>
</tr>
</thead>
<tbody>
<tr>
<td>o₁</td>
<td>short</td>
<td>blond</td>
<td>blue</td>
<td>+</td>
</tr>
<tr>
<td>o₂</td>
<td>short</td>
<td>blond</td>
<td>brown</td>
<td>-</td>
</tr>
<tr>
<td>o₃</td>
<td>tall</td>
<td>red</td>
<td>blue</td>
<td>+</td>
</tr>
<tr>
<td>o₄</td>
<td>tall</td>
<td>dark</td>
<td>blue</td>
<td>-</td>
</tr>
<tr>
<td>o₅</td>
<td>tall</td>
<td>dark</td>
<td>blue</td>
<td>-</td>
</tr>
<tr>
<td>o₆</td>
<td>tall</td>
<td>blond</td>
<td>blue</td>
<td>+</td>
</tr>
<tr>
<td>o₇</td>
<td>tall</td>
<td>dark</td>
<td>brown</td>
<td>-</td>
</tr>
<tr>
<td>o₈</td>
<td>short</td>
<td>blond</td>
<td>brown</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 2: An example of partition by ID3

The Universe
{ o₁, o₂, o₃, o₄, o₅, o₆, o₇, o₈ }

- hair-blonde
  -{ o₁, o₂, o₆, o₈ }
  +/{ o₁, o₆ }
  -/{ o₂, o₈ }

- hair-darker
  -{ o₇, o₄, o₅ }

- hair-red
  +{ o₅ }

- eyes-blue
  -{ o₁, o₆ }

- eyes-brown
  +{ o₈ }
  -{ o₂ }

34
PRISM (for covering)

- ID3: using the notion of entropy for partitioning
- PRISM: using the conditional probability for covering
  - Conditional probability: \( P(A|B) \)
    - The possible of A when B happens

<table>
<thead>
<tr>
<th>object</th>
<th>height</th>
<th>hair</th>
<th>eyes</th>
<th>class</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>

\[
P(+) | hair=red = \frac{|\{o_1, o_3, o_6\}|}{|\{o_1, o_3, o_4, o_5, o_6\}|} = 1
\]
\[
P(+) | hair=blond = \frac{|\{o_1, o_6\}|}{|\{o_1, o_3, o_4, o_5, o_6\}|} = 0.5
\]
Extraction of classification rules using Granules

- **ID3**: using only an attribute for each step
- **PRISM**: executing rule extraction procedures for each class
- **Extended one**: using various attributes for each step and executing rule extraction procedure at once.

<table>
<thead>
<tr>
<th>object</th>
<th>height</th>
<th>hair</th>
<th>eyes</th>
<th>class</th>
</tr>
</thead>
<tbody>
<tr>
<td>o₁</td>
<td>short</td>
<td>blond</td>
<td>blue</td>
<td>+</td>
</tr>
<tr>
<td>o₂</td>
<td>short</td>
<td>blond</td>
<td>brown</td>
<td>-</td>
</tr>
<tr>
<td>o₃</td>
<td>tall</td>
<td>red</td>
<td>blue</td>
<td>+</td>
</tr>
<tr>
<td>o₄</td>
<td>tall</td>
<td>dark</td>
<td>blue</td>
<td>-</td>
</tr>
<tr>
<td>o₅</td>
<td>tall</td>
<td>dark</td>
<td>blue</td>
<td>-</td>
</tr>
<tr>
<td>o₆</td>
<td>tall</td>
<td>blond</td>
<td>blue</td>
<td>+</td>
</tr>
<tr>
<td>o₇</td>
<td>tall</td>
<td>dark</td>
<td>brown</td>
<td>-</td>
</tr>
<tr>
<td>o₈</td>
<td>short</td>
<td>blond</td>
<td>brown</td>
<td>-</td>
</tr>
</tbody>
</table>

The Universe
{₀₁,₀₂,₀₃,₀₄,₀₅,₀₆,₀₇,₀₈}

- **hair-dark** {₀₄,₀₅,₀₇} -
- **eyes-brown** {₀₂,₀₇,₀₈} -
- **hair-red** {₀₃} +
- **hair-blond** {₀₁,₀₂,₀₆,₀₈} +/-
- **eyes-blue** {₀₁,₀₆} +
- redundant granule
Q n A