TITCH:
Attribute selection based on discrimination power and frequency

Philipp Spanger, Kurosawa Takahiro, Tokunaga Takenobu
Department of Computer Science
Tokyo Institute of Technology
Two factors in attribute selection

- Human’s general preference on object attributes e.g. colour > orientation → Frequency of attributes
- Salience of attributes in a specific case (particular situation in a domain) → Discrimination power
**Base algorithm**  
(Using case specific attribute salience)

- Ranking attribute-value pairs according to their discrimination power
  - Disc. power = No. of objects excluded by specifying an attr.-value pair
- Selecting attributes one-by-one until unique identification of the target

<table>
<thead>
<tr>
<th>Domain</th>
<th>Dice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>static</td>
</tr>
<tr>
<td>Furniture</td>
<td>0.588</td>
</tr>
<tr>
<td>People</td>
<td>0.559</td>
</tr>
</tbody>
</table>

©2007 Tokunaga Takenobu
Improvement of base algorithm (Introducing Human’s general preference on attr.)

- Human’s general preference on attributes
  \( \equiv \) Frequency of attributes in the domain

- Assumption: Salient attributes tend to be used frequently

- Weighting attributes based on frequency of occurrence in the data (absolute, relative)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Dice Base</th>
<th>Dice Abs.</th>
<th>Dice Rel.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Furniture</td>
<td>0.588</td>
<td>0.685</td>
<td>0.707</td>
</tr>
<tr>
<td>People</td>
<td>0.559</td>
<td>0.651</td>
<td>0.648</td>
</tr>
<tr>
<td>People+</td>
<td>---</td>
<td>0.683</td>
<td>0.678</td>
</tr>
</tbody>
</table>
Indispensable features

• “Type”
  • Used all instances in both domains
• Dependency between attributes
  • hairColour → hasHair, hasBeard
• What’s common in above two cases?
  • Those could be realised as a head of NP
Difficulty of cases

- Metric representing a difficulty of a given case?
- No. of possible different sets of attr.-value pairs uniquely identifying the target
- But, no correlation the success of our algorithm
- This reflects the fact that humans select attributes from a very limited set of combinations, independent of the search space