Running example

An electric guitar and bass player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.

1. Fish sense
2. Musical sense
3. ...

Senseval example

<instance id="add.v.bnc.00003517" docsrc="BNC">
<answer instance="add.v.bnc.00003517" senseid="42603"/>
<context>
But this one looks impeccable. There's only one solution, and that is to destroy the picture. It's second rate, in any case (he added smiling), so the artistic loss to the world will be nil. As for the purchaser, tell him to come to my house and choose another picture something that both he and I like.
SNAPSHOTS: THE ARTICLES
</context>
</instance>
Senseval answer file

add.v add.v.bnc.00000134 42601 42606
add.v add.v.bnc.00000242 42603
add.v add.v.bnc.00000837 42601 42606
add.v add.v.bnc.00001199 42601
add.v add.v.bnc.00001354 42606
add.v add.v.bnc.00002656 42603
add.v add.v.bnc.00002929 42601
add.v add.v.bnc.00003185 42601 42606
add.v add.v.bnc.00003517 42603
add.v add.v.bnc.00004108 42601

WEKA: the bird

WEKA: the software

- Machine learning/data mining software written in Java (distributed under the GNU Public License)
- Used for research, education, and applications
- Complements “Data Mining” by Witten & Frank
- Main features:
  - Comprehensive set of data pre-processing tools, learning algorithms and evaluation methods
  - Graphical user interfaces (incl. data visualization)
  - Environment for comparing learning algorithms

WEKA only deals with “flat” files

@relation heart-disease-simplified

@attribute age numeric
@attribute sex { female, male}
@attribute chest_pain_type { typ_angina, asympt, non_anginal, atyp_angina}
@attribute cholesterol numeric
@attribute exercise_induced_angina { no, yes}
@attribute class { present, not_present}

@data
63,male,typ_angina,233,no,not_present
67,male,asympt,286,yes,present
67,male,asympt,229,yes,present
38,female,non_anginal,?,no,not_present
...

Copyright: Martin Kramer (mkramer@wxs.nl)
### WEKA only deals with “flat” files

<table>
<thead>
<tr>
<th>@relation heart-disease-simplified</th>
</tr>
</thead>
<tbody>
<tr>
<td>@attribute age numeric</td>
</tr>
<tr>
<td>@attribute sex { female, male }</td>
</tr>
<tr>
<td>@attribute chest_pain_type { typ_angina, asympt, non_anginal, atyp_angina }</td>
</tr>
<tr>
<td>@attribute cholesterol numeric</td>
</tr>
<tr>
<td>@attribute exercise_induced_angina { no, yes }</td>
</tr>
<tr>
<td>@attribute class { present, not_present }</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>@data</th>
</tr>
</thead>
<tbody>
<tr>
<td>63, male, typ_angina, 233, no, not_present</td>
</tr>
<tr>
<td>67, male, asympt, 286, yes, present</td>
</tr>
<tr>
<td>67, male, asympt, 229, yes, present</td>
</tr>
<tr>
<td>38, female, non_anginal, ?, no, not_present</td>
</tr>
</tbody>
</table>

### WEKA demo

### Collocational features

- Encode information about the lexical inhabitants of specific positions located to the left or right of the target word.
- E.g. the word, its root form, its part-of-speech
  - An electric guitar and bass player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.

### Co-occurrence features

- Encodes information about neighboring words, ignoring exact positions.
  - Select a small number of frequently used content words for use as features
    - 12 most frequent content words from a collection of bass sentences drawn from the WSJ: fishing, big, sound, player, fly, rod, pound, double, runs, playing, guitar, band
  - Co-occurrence vector (window of size 10)
  - Attributes: the words themselves (or their roots)
  - Values: number of times the word occurs in a region surrounding the target word

<table>
<thead>
<tr>
<th>pre2-word</th>
<th>pre2-pos</th>
<th>pre1-word</th>
<th>pre1-pos</th>
<th>fol1-word</th>
<th>fol1-pos</th>
<th>fol2-word</th>
<th>fol2-pos</th>
</tr>
</thead>
<tbody>
<tr>
<td>guitar</td>
<td>NN1</td>
<td>and</td>
<td>CJC</td>
<td>player</td>
<td>NN1</td>
<td>stand</td>
<td>VVB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Inductive ML framework

Examples of task
(features + class)

ML Algorithm

Novel example
(features)

Classifier
(program)

correct word sense
description of context

learn one such classifier for each lexeme to be disambiguated

class