Supervised Learning from Corpora

Overview
1. Transformation-Based Learning
2. The Back-Off Method
3. A combination of supervised and unsupervised results

Training material
- the NEGRA test set (6064 test cases)
- the CZ test set (4562 test cases)
- as basis for small training sets!!

The sparse data problem
Many quadruples will occur rarely!
Therefore: clustering is needed
- verbs → lemmas
- contracted prepositions → base forms
- proper names → class labels
- numbers → number tag
- nouns → lemmas (of last compound element)

Transformation-Based Learning
- developed by Eric Brill for Part-of-Speech Tagging
- idea: learn transformation rules from the manually disambiguated cases based on rule templates

Transformation-Based Learning for PP-attachment
- idea: learn transformation rules from the manually disambiguated cases based on rule templates
  - start with "noun attachment"
  - in each step determine the rule that contributes most to the correction of the training set
  - the rule templates use the quadruple \((V, N1, P, N2)\) and can access
    - one specific word (4 templates) or
    - any combination of two words (6 templates) or
    - any triple that contains the preposition (3 templates)
Examples for learned rules

- change from noun att. to verb att. if N1 = <person>
- change from noun att. to verb att. if P = auf
- change from verb att. to noun att. if N1 = <Person>
  && P = von

- Rules learned from CZ test set and NEGRA test set are very similar!
- BUT: Rule learning even for our small training corpus takes hours!!

Application of the Rules

- Start with the default attachment decision and correct with the transformation rules.

Disambiguation Results for the Transformation-Based Method

<table>
<thead>
<tr>
<th>Training corpus</th>
<th>Test set</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEGRA</td>
<td>CZ</td>
<td>72.3%</td>
</tr>
<tr>
<td>NEGRA + 4/5 CZ</td>
<td>1/5 CZ * 5</td>
<td>Ø = 76.4%</td>
</tr>
</tbody>
</table>

Back-off Method

- by Collins and Brooks
- Idea: learn attachment tendencies from manually disambiguated cases
- in case of missing information back-off to the next level
  - quadruples
  - triples (which include the preposition)
  - pairs (which include the preposition)
  - prepositions alone

Back-off Algorithm

```
if (freq(V,N1,P,N2) > 0) then
  if (freq(noun_att, V,N1,P,N2) / (freq(V,N1,P,N2) ) > 0.5
    then noun attachment
    else verb attachment
  elsif  (( freq(V,N1,P) + freq(V,P,N2) + freq(N1,P,N2)) > 0)
    then noun attachment
    else verb attachment
  elsif  (( freq(V,P) + freq(P,N2) + freq(N1,P) ) > 0 ) then
    ...
```

Back-off Example

```
V: geben  N1: Firma  P: in  N2: <GEO>  Fct:MNR
freq(n_att, V,N1,P): 1
freq(n_att, V,P,N2):  5
freq(n_att, N1,P,N2): Sum  6
freq(V,N1,P):  1
freq(V,P,N2):  5
freq(N1,P,N2):  Sum  5
Ratio 0.2
```
Results for the Back-off method

<table>
<thead>
<tr>
<th>decision level</th>
<th># of cases</th>
<th>accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>quadruples</td>
<td>8</td>
<td>100.00%</td>
</tr>
<tr>
<td>triples</td>
<td>329</td>
<td>88.75%</td>
</tr>
<tr>
<td>pairs</td>
<td>3040</td>
<td>75.66%</td>
</tr>
<tr>
<td>preposition</td>
<td>1078</td>
<td>64.66%</td>
</tr>
<tr>
<td>default</td>
<td>14</td>
<td>64.29%</td>
</tr>
<tr>
<td>total</td>
<td>4469</td>
<td>73.98%</td>
</tr>
<tr>
<td>(coverage: 100%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results for the Back-off Method

<table>
<thead>
<tr>
<th>Training corpus</th>
<th>Test set</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEGRA</td>
<td>CZ</td>
<td>74.0%</td>
</tr>
<tr>
<td>CZ</td>
<td>NEGRA</td>
<td>68.3%</td>
</tr>
<tr>
<td>NEGRA + 4/5 CZ</td>
<td>1/5 CZ * 5</td>
<td>79.4%</td>
</tr>
</tbody>
</table>

Intertwined Combination

1. Support verb unit
2. -
3. -
4. Triple comparison
5. Pair comparison
6. Threshold compar.
7. -
8. -
9. -

Unsupervised

1. Support verb unit
2. -
3. -
4. Triple comparison
5. Pair comparison
6. Threshold compar.
7. -
8. -
9. -

Supervised

1. -
2. Quadruples
3. Triples
4. -
5. -
6. -
7. Pairs
8. Preposition
9. Default

Comparison of results

Unsupervised: 79.14%
Supervised: 73.98%
Combined: 80.98%

Conclusions

• Supervised methods lead to better results than unsupervised methods given enough treebank material from the right domain.
• Unsupervised method is as good as supervised method over small training corpus.
• Combination of unsupervised and supervised leads to the best results.
Conclusions

- Beware: V,N1,P,N2 makes the PP attachment task look easier than it is!!
- It reduces the PP attachment task to a simple case.
- But often: sequences of NPs and PPs:
  - V,NP_PP_PP or
  - V,NP,NP_PP