Exploring Lexicalized Features for Coreference Resolution

Anders Björkelund and Pierre Nugues

June 24, 2011
Overview

- Pair-wise classifier based on Soon et al. (2001)
- Syntactic dependencies obtained through an automatic conversion from the constituents
- Large number of lexical and dependency-based feature templates
- Automatic feature selection
System Architecture

- **Preprocessing**
  - Mention extraction — All NPs and possessive pronouns
  - Conversion to syntactic dependencies using the LTH converter

- **Pair-wise classifier using logistic regression (LIBLINEAR)**
  - Closest-first clustering for pronouns
  - Best-first clustering for nonpronominals

- **Postprocessing (next slide)**
  - Recovery of missed mentions using string matching
Postprocessing

- Not all mentions are extracted during mention extraction
  - The automatically parsed constituents contain mistakes
  - NML constituents were disregarded during mention extraction
- Obvious and easy examples include proper nouns
- Recovering missed mentions:
  - Search the document for spans of one or more proper nouns whose immediate parent was not clustered
  - Try to match this span of proper nouns to all mentions that were clustered by the classifier using string match
  - If match, add this span to corresponding chain
- Example
  
  (NP (NML (NNP Hong) (NNP Kong)) (NN cinema))
Features (baseline)

- Baseline system: Reimplementation of the Soon et al. (2001) system with 12 features, e.g.
  - StringMatch
  - GenderAgreement
  - AnaphorIsPronoun
  - AnaphorIsDefinite
  - ...

- These features are extracted using hand-crafted rules.
- They can often be simply reframed in terms of dependencies:
  - IsPronoun can be deduced from POS tag of head word
  - IsDefinite can be deduced from surface form of leftmost child of head word
To enable a systematic search without requiring prior knowledge, we defined additional feature templates. Using the dependency graph of the noun phrase:

- Surface form, POS tag, dependency label of HeadWord, LeftMostChild, RightMostChild, HeadGovernor, HeadLeftSibling, HeadRightSibling
- Dependency graph paths, i.e. direction of edges and Form, POS, or dependency label

A number of variations of semantic role features

Total of ca. 60 feature templates

(See paper for details)
Feature Selection

- Baseline set was the Soon et al (2001) feature set
- Pool of feature templates including all above and a set of manually selected pairs, e.g.
  - AntecedentHeadForm + AnaphorHeadForm
  - AntecedentHeadLeftMostChild + AnaphorHeadLeftMostChild
- Greedy forward-backward selection, incrementally adding or removing one feature template from the current set
- Cross-validated over the training set, in order not to skew it towards the development set
- Optimized for the CoNLL score
Impact of the postprocessing step:

<table>
<thead>
<tr>
<th></th>
<th>MD</th>
<th>MUC</th>
<th>BCUB</th>
<th>CEAFM</th>
<th>CEAFE</th>
<th>BLANC</th>
</tr>
</thead>
<tbody>
<tr>
<td>No PP</td>
<td>66.56</td>
<td>54.61</td>
<td>65.93</td>
<td>51.91</td>
<td>40.46</td>
<td>69.36</td>
</tr>
<tr>
<td>With PP</td>
<td>67.21</td>
<td>55.62</td>
<td>66.29</td>
<td>52.51</td>
<td>40.67</td>
<td>70.00</td>
</tr>
<tr>
<td>Increase</td>
<td>0.65</td>
<td>1.01</td>
<td>0.36</td>
<td>0.60</td>
<td>0.21</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Overall beneficial – increased precision and recall across all metrics
### Results (evaluation set)

- Results on the test set – Fourth place in the Shared Task

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>P</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mention detection</td>
<td>69.87</td>
<td>68.08</td>
<td>68.96</td>
</tr>
<tr>
<td>MUC</td>
<td>60.20</td>
<td>57.10</td>
<td>58.61</td>
</tr>
<tr>
<td>BCUB</td>
<td>66.74</td>
<td>64.23</td>
<td>65.46</td>
</tr>
<tr>
<td>CEAFM</td>
<td>51.45</td>
<td>51.45</td>
<td>51.45</td>
</tr>
<tr>
<td>CEAFE</td>
<td>38.09</td>
<td>41.06</td>
<td>39.52</td>
</tr>
<tr>
<td>BLANC</td>
<td>71.99</td>
<td>70.31</td>
<td>71.11</td>
</tr>
<tr>
<td>Official CoNLL score</td>
<td>55.01</td>
<td>54.13</td>
<td>54.53</td>
</tr>
</tbody>
</table>

- Our system makes no use of global optimization or constraints
- We believe feature selection was a key ingredient
- This technique should be replicable to other languages
Questions

- Questions?