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

Feature Templates

- 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

Postprocessing (development set)

- Impact of the postprocessing step:

	MD	MUC	BCUB	CEAFM	CEAFE	BLANC
No PP	66.56	54.61	65.93	51.91	40.46	69.36
With PP	67.21	55.62	66.29	52.51	40.67	70.00
Increase	0.65	1.01	0.36	0.60	0.21	0.64

- Overall beneficial – increased precision and recall across all metrics

Results (evaluation set)

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

	R	P	F1
Mention detection	69.87	68.08	68.96
MUC	60.20	57.10	58.61
BCUB	66.74	64.23	65.46
CEAFM	51.45	51.45	51.45
CEAFE	38.09	41.06	39.52
BLANC	71.99	70.31	71.11
Official CoNLL score	55.01	54.13	54.53

- 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?