ElkfEd/IDC alias BART

Yannick Versley Simone Ponzetto

Jason Smith Vladimir Eidelman Alan Jern

Massimo Poesio Xiaofeng Yang

Introduction

- Modularized and revamped version of EML-R system
- Full coreference including mention detection
- Will be publically available as Open Source
- Useful Java-only subset (e.g. for teaching)

Components

- Preprocessing
  aggregate data in MMAX2 annotation layers
- Mention extraction
  create markables from chunks/NEs
- Extract information about mentions
  mention type, semantic class ...
- Encode coref into classifier decisions and extract features

Components

Modular pipeline architecture
Standoff annotation

- Preprocessing
  aggregate data in MMAX2 annotation layers
- Mention extraction
  create markables from chunks/NEs
- Extract information about mentions
  mention type, semantic class ...
- Encode coref into classifier decisions and extract features

Components

- Preprocessing
  aggregate data in MMAX2 annotation layers
- Mention extraction
  create markables from chunks/NEs
- Extract information about mentions
  mention type, semantic class ...
- Encode coref into classifier decisions and extract features

Feature set and learners
(and learner settings!)
described in XML file

Machine Learning Infrastructure

- WEKA Machine Learning Toolkit
  - C4.5, RIPPER, other learning modules
- SVMlight-TK
  - SVMs, different kernels (linear, polynomial, etc.)
  - Tree kernels
  - Custom kernels
- MaxEnt-based ranking
  - find best among a set of candidates

Flexibility in Preprocessing

- Chunker (YamCha) vs. Parser (Charniak)
- Stanford NER vs. Carafe
- Mentions from Chunks/NEs vs.
  Mentions from Mention Tagger (Carafe)

Resolution Algorithms

- Closest-first decoding (Soon et al., 2001)
- Separate classifiers for pronouns/non-pronouns
- Ranking-based resolution
- Stacking ranking+classifier
- Global models

Quantitative / Qualitative Evaluation

- MUC scoring (per document, in total)
- Link-based scores, by type of anaphora *(pronouns, appositions / copula, names, nominals)*
- Qualitative evaluation *inspect results in MMAX2*