

EIkfEd/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

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

Modular pipeline architecture
Standoff annotation

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*

```
MMAX2 1.12 /home/yannick/tmp/MUC-MMAX/muc6/test/ws93_022.0297.mmax [modified]
File Settings Display Tools Plugins Info ShowMLPanel
DOCID: wsj93_022_0297
DOCNO: 930504-0023 .
HL: @IBM@ appoints @ @Chrysler@ 's york@ as @finance@ chief@ --- @ @computer maker@ 's move@ signals
strategy of cuts @ in @its@ costs , asset sales @ ---- @ by @Michael W Miller@ and @Douglas Lavin@ @ staff
reporters of the wall street journal
DD: 05/04/93
wall street journal ( j ) , page 33 @ IBM@ automobiles ( aut ) , computers ( cpr )
TXT: @International Business Machines Corp.@ continued @its@ executive makeover by hiring @Jerome B. York@
an architect of the turnaround at @Chrysler Corp.@ , to become @chief financial officer@ .
@Mr. York@ , 54 years old , is a @West Point@ graduate who helped transform @Chrysler@ by slashing @costs@ and
selling billions of dollars in assets .
@His@ appointment is a strong sign that @IBM@ 's new chairman , @Louis V. Gerstner Jr.@ , plans a similar
strategy at @the wounded computer giant@ .
@Mr. Gerstner@ raced to hire @Mr. York@ after meeting @him@ for the first time just three weeks ago in @IBM@ 's
Manhattan offices .
In @his@ first month , @Mr. Gerstner@ has also brought in outsiders to run @IBM@ 's communications and
disk-drive business , and is searching for a new head of personnel .
@Mr. York@ was @executive vice president for @finance@ and a board member at @Chrysler@ , where @he@ spent
12 years in financial posts and running several @car@ and truck divisions .
@Chrysler@ did n't name a successor @yesterday@ .
```