Coreference Resolution

Different German & English Coreference Resolution Models for Multi-Domain Content Curation Scenarios

Ankit Srivastava, Sabine Weber, Peter Bourgonje, and Georg Rehm
DFKI GmbH – Forschungsbereich Sprachtechnologie, Berlin

GSCL 2017, 14th September, Berlin
About this Presentation

- Introduction to Coreference Resolution
- Coreference Resolution for English
- Coreference Resolution for German
- Our Approaches: $\text{Coref}_{\text{rule}}$, $\text{Coref}_{\text{stat}}$, $\text{Coref}_{\text{proj}}$
- Coreference Resolution for Digital Curation
- Endpoint
COREFERENCE RESOLUTION
Audi is an automaker that makes luxury cars and SUVs. The company was born in Germany.

It was established by August Horch in 1910. Horch had previously founded another company and his models were quite popular. Audi started with four cylinder models. By 1914, Horch’s new cars were racing and winning.

August Horch left the Audi company in 1920 to take a position as an industry representative for the German motor vehicle industry federation.

Currently Audi is a subsidiary of the Volkswagen group and produces cars of outstanding quality.

Source: Coreference Resolution presentation by Shumin Wu and Nicolas Nicolov of J.D. Power and Associates
What is Coreference Resolution?

• Process of identifying all words & phrases in a document that refer to the same entity

• Core of Natural Language Understanding (NLU) since 1960s

• Documents usually contain the full named entity once or a few times. For full NLU, coreference resolution is essential

• Can be meaningfully applied in Question Answering, Named Entity Recognition, Machine Translation, Summarisation
Coreference & Co.

Anaphora
*The music was so loud it couldn’t be enjoyed.*

Cataphora
*Despite her difficulty, Wilma came to understand the point.*

Split antecedents
*Carol told Bob to attend the party. They arrived together.*

Coreferring noun phrases
*Some of our colleagues will help us. These people will earn our trust.*
Motivation

• Digital Curation Platform for Knowledge Workers
  – Named Entity Recognition, Entity & Relation Extraction, Translation, Summarisation, Timelining, Clustering

• Benefit from Coreference: Disambiguation
  – Explore tools available for both German & English
  – Build our own… evaluate
SUMMARY OF APPROACHES
Approaches to Coreference

• 3 Paradigms
  – Rule-based (Heuristics)
  – Machine Learning (Mention-Rank Model)
  – Knowledge-based (Crosslingual Projections)

• Coreference Resolution for English
  – [Raghunathan et al., 2010] [Clark & Manning, 2015 I 2016]

• Coreference Resolution for German
  – [Verseley et al., 2008] [Krug at al., 2015] [Roesiger & Riester, 2015] [Tuggener, 2016]
Evaluation of Coreference

• Benchmarking Shared Tasks on Standard datasets
  – Message Understanding Conference (MUC)
  – Automatic Content Extraction (ACE)
  – Computational Natural Language Learning (CoNLL)
  – Semantic Evaluation (SemEval)
  – Coreference Resolution beyond OntoNotes (CORBON)

• Evaluation Metrics
  – Message Understanding Conference F-measure (MUC6)
  – Bagga, Baldwin, Biermann (B³)
  – Constrained Entity-Alignment F-Measure (CEAF)
Evaluation: MUC-6 F-Measure

Reference:

System output:

Count the number of corresponding links between mentions

Precision = 4/5
Recall = 4/6
F-measure = 2 * Precision * Recall / (Precision + Recall) = 0.727
Evaluation Metrics Summary

• MUC6 F-measure
  – Ignores single mention entities
  – Potentially biased toward large clusters
  – No one-to-one entity mapping guarantee

• $B^3$
  – Set view of mentions in an entity
  – Based on number of corresponding mentions between entities averaged over total number of mentions
  – Does not provide one-to-one entity mapping

• CEAF
  – One-to-one entity mapping
  – Optimal mapping can be tuned to a different similarity measure
THREE IMPLEMENTATIONS

Coref$_{rule}$  Coref$_{stat}$  Coref$_{proj}$
Rule-based (Multi-Sieve)

• English version Based on Stanford CoreNLP

• German version: in-house implementation

• Idea of an annotation pipeline
  – Sentence splitting, tokenisation, parsing, morphology
Multi-Sieve Coreference

- Sieve 7: German Specific Processing
- Sieve 6: Named Entity Recognition
- Sieves 3-5: NP Head Matching
- Sieve 2: Precise Constructs
- Sieve 1: Exact Match
Multi-Sieve Coreference

- Parse a document
- Get Noun Phrases, Pronominal Phrases
- Cluster them via sieve heuristics
- **Exact Match**: If NPs match each other in a context window of 5 (with stemming), then link them
- “Der Hund” / “Des Hundes”

Sieve 1: Exact Match

Sieve 2: Precise Constructs

Sieves 3-5: NP Head Matching

Sieve 6: Named Entity Recognition

Sieve 7: German Specific Processing
Multi-Sieve Coreference

- Models linked if appositive or predicative nominative constructs are detected
- Appositive: "Donald Trump, President of USA"
Multi-Sieve Coreference

- If head word of two NPs is same, they are coreferent
- Some Relaxations and Rules
Examples of Match Heuristics

• Compatible mentions:
  – Exact string match of capitalized mentions
    “Trump” & “Trump”
  – Exact string match of mentions within a sentence
    “car” & “car”
  – Acronyms
    “USA” & “United States of America”
  – First person / Second person / Third person pronoun
    “I” & “me”, “you” & “yours”, “he” & “him”, “she” & “her”

• Incompatible mentions:
  – Different acronyms
    “USA” & “UK”
  – Personal, gender, number disagreement
    “I” & “you”, “he” & “she”, “car” & “cars”
Evaluation: Sieve Settings

1. All Sieves in place
2. All mentions but no coreference links
3. \{1\} after deletion of clusters with no mentions
4. \{1\} with insertion of clusters with no mentions added last

<table>
<thead>
<tr>
<th>System</th>
<th>MUC</th>
<th>B-Cube</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting 1</td>
<td>54.4</td>
<td>11.2</td>
</tr>
<tr>
<td>Setting 2</td>
<td>70.5</td>
<td>23.1</td>
</tr>
<tr>
<td>Setting 3</td>
<td>58.9</td>
<td>15.0</td>
</tr>
<tr>
<td>Setting 4</td>
<td>56.1</td>
<td>12.0</td>
</tr>
</tbody>
</table>
Statistical (Mention-Ranking)

• Based on Stanford CoreNLP (English & German)
  – English trained and evaluated on CoNLL ’11, ‘12
  – German trained on TüBa/D-Z, evaluated on SemEval ’10

• 4 Types of Features learned using Dagger [Ross 2011]
  – Distance
  – Syntactic
  – Semantic
  – Lexical

• Issues in out-of-domain adaptation
Features

- Distance features: the distance between the two mentions in a sentence, number of mentions

- Syntactic features: number of embedded NPs under a mention, Part-Of-Speech tags of the first, last, and head word (based on the German parsing models included in the Stanford CoreNLP (Rafferty and Manning, 2008)

- Semantic features: named entity type, speaker identification

- Lexical Features: the first, last, and head word of the current mention
Projection (Crosslingual)

- Coreference for German based on English models
- Transferring Models Vs Transferring Data
- Corbon 2017 English—German Data
## Comparative Evaluation

<table>
<thead>
<tr>
<th>System</th>
<th>MUC</th>
<th>B-Cube</th>
</tr>
</thead>
<tbody>
<tr>
<td>BART</td>
<td>45.3</td>
<td>64.5</td>
</tr>
<tr>
<td>CorZu</td>
<td>60.1</td>
<td>58.9</td>
</tr>
<tr>
<td>Sieve</td>
<td>49.2</td>
<td>45.3</td>
</tr>
<tr>
<td>Statistical</td>
<td>56.3</td>
<td>50.4</td>
</tr>
<tr>
<td>Neural</td>
<td>60.0</td>
<td>56.8</td>
</tr>
</tbody>
</table>

**English: CoNLL 2012**

<table>
<thead>
<tr>
<th>System</th>
<th>MUC</th>
<th>B-Cube</th>
</tr>
</thead>
<tbody>
<tr>
<td>CoRefGer-rule</td>
<td>50.2</td>
<td>63.3</td>
</tr>
<tr>
<td>CoRefGer-stat</td>
<td>40.1</td>
<td>45.3</td>
</tr>
<tr>
<td>CoRefGer-proj</td>
<td>35.9</td>
<td>40.3</td>
</tr>
</tbody>
</table>

**German: SemEval 2010**

**Deutsches Forschungszentrum für Künstliche Intelligenz GmbH**

GSCL 2017 - Coreference for Digital Curation
MULTI DOMAIN CONTENT CURATION SCENARIOS
Curation Case Studies

<table>
<thead>
<tr>
<th>Corpora</th>
<th>Language</th>
<th>Documents</th>
<th>Words</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mendelsohn</td>
<td>DE</td>
<td>2,501</td>
<td>699,213</td>
<td>Personal letters</td>
</tr>
<tr>
<td>Mendelsohn</td>
<td>EN</td>
<td>295</td>
<td>21,226</td>
<td>Personal letters</td>
</tr>
<tr>
<td>Vikings</td>
<td>EN</td>
<td>12</td>
<td>298,577</td>
<td>Wikipedia and E-books</td>
</tr>
<tr>
<td>News</td>
<td>DE</td>
<td>1,037</td>
<td>716,885</td>
<td>News articles and summaries</td>
</tr>
</tbody>
</table>
Coreference for Curation

- Applied English & German coreference models on different datasets
- $\text{Coref}_{\text{rule}}$ outperforms $\text{Coref}_{\text{stat}}$, $\text{Coref}_{\text{proj}}$ in terms of number of mentions

<table>
<thead>
<tr>
<th>Dataset</th>
<th>SENTS.</th>
<th>WORDS</th>
<th>MENTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mendelsohn EN</td>
<td>21K</td>
<td>109K</td>
<td>48%</td>
</tr>
<tr>
<td>Mendelsohn DE</td>
<td>34K</td>
<td>681K</td>
<td>26%</td>
</tr>
<tr>
<td>Vikings EN</td>
<td>39K</td>
<td>310K</td>
<td>49%</td>
</tr>
<tr>
<td>News Stories DE</td>
<td>53K</td>
<td>369K</td>
<td>25%</td>
</tr>
</tbody>
</table>
About this Presentation

• Introduction to Coreference Resolution
• Coreference Resolution for English
• Coreference Resolution for German
• Our Approaches: Coref\textsubscript{rule}, Coref\textsubscript{stat}, Coref\textsubscript{proj}
• Coreference Resolution for Digital Curation

• Endpoint
Conclusions

• Performed Coreference Resolution
  – In both English & German
  – On a variety of text types
  – For competing approaches (sieve, mention-rank, projection)

• Successful in coreference resolution for curation datasets such as an archive of letters, research materials for exhibition, news articles & downstream applications

• Currently, best choice is Multi Sieve (Rule-based) approach for out-of-domain processing
Thank You!

Email: ankit.srivastava@dfki.de
EXTRA SLIDES