Resolving Entity Coreference in Croatian with a Constrained Mention-Pair Model

Goran Glavaš and Jan Šnajder
TakeLab UNIZG

BSNLP 2015 @ RANLP, Hissar

10 Sep 2015

Background & Motivation

- Entity coreference resolution (CR)
  - Identifying **different mentions** of the **same entity**
  - Important NLP task with **numerous applications**: relation extraction, question answering, summarization, ...
- Easy to define but difficult to tackle
  - **External knowledge** often required (e.g., *"U.S. President"* $\Leftrightarrow$ *"Barack Obama"*)

Existing Work

- Early, **rule-based** CR focused on theories of discourse such as *focusing* and *centering* (Sidner 1979; Grosz et al., 1983)
- Shift to **machine-learning** approaches occurred with appearance of manually annotated coreference data (MUC)

Existing Work

- Early, **rule-based** CR focused on theories of discourse such as *focusing* and *centering* (Sidner 1979; Grosz et al., 1983)
- Shift to **machine-learning** approaches occurred with appearance of manually annotated coreference data (MUC)
- The **mention-pair model** is the most widely applied coreference resolution model (Aone and Bennett, 1995)
  - A **binary classifier** for pairs of event mentions
  - **Fails to account for transitivity** of the coreference relation

Existing Work

- Early, **rule-based** CR focused on theories of discourse such as *focusing* and *centering* (Sidner 1979; Grosz et al., 1983)
- Shift to **machine-learning** approaches occurred with appearance of manually annotated coreference data (MUC)
- The **mention-pair model** is the most widely applied coreference resolution model (Aone and Bennett, 1995)
  - A **binary classifier** for pairs of event mentions
  - **Fails to account for transitivity** of the coreference relation
- More complex models failed to significantly outperform the mention-pair model
  - Entity-mention models (Daume III and Marcu, 2005)
  - Ranking models (Yang et al., 2008)

Existing Work

- Besides large body of work for English, much work has been done for other major languages as well
  - Spanish (Palomar et al., 2001; Sapena et al., 2010)
  - Italian (Kobdani and Schütze 2010; Poesio et al., 2010)
  - German (Versley, 2006; Wunsch, 2010)
  - Chinese (Converse, 2006; Kong and Zhou, 2010)
  - ...

Existing Work

- Besides large body of work for English, much work has been done for other major languages as well
  - Spanish (Palomar et al., 2001; Sapena et al., 2010)
  - Italian (Kobdani and Schütze 2010; Poesio et al., 2010)
  - German (Versley, 2006; Wunsch, 2010)
  - Chinese (Converse, 2006; Kong and Zhou, 2010)
  - ...
- Research for **Slavic languages** has been quite limited
  - Substantial research for Polish (Marciniak, 2002; Matysiak, 2007; Kopec and Ogrodniczuk, 2012)
  - Czech (Linh et al., 2009)
  - Bulgarian
    (Zhikov et al., 2013)

Coreference Resolution for Croatian

Data Annotation

- We adopt the CR type scheme for Polish (Ogrodniczuk et al., 2013)

CR type	Example
IDENTITY	**Premijer** *je izjavio da* **on** *nije odobrio taj zahtjev.* (**The Prime Minister** said **he** didn't grant that request.)
HYPER-HYPO	*Ivan je kupio novi* **automobil**. **Taj Mercedes** *je čudo od auta.* (Ivan bought a new **car**. **That Mercedes** is an amazing car.)
MERONYMY	*Od* **jedanaestorice rukometaša** *danas je igralo samo njih* **osam**. (Only **eight** out of **eleven handball players** played today.)
METONYMY	**Dinamo** *je jučer pobijedio Cibaliju.* **Zagrepčani** *su postigli tri pogotka.* (**Dinamo** defeated Cibalia yesterday. **Zagreb boys** scored three goals.)
∅-ANAPHORA	**Marko** *je išao u trgovinu.* **Kupio** *je banane.* (**Marko** went to the store. [He] **bought** bananas.)

Data Annotation

- News articles corpus of 285 documents
- Six trained annotators
  - Detailed annotation guidelines
  - In-house developed annotation tool

Data Annotation

- News articles corpus of 285 documents
- Six trained annotators
  - Detailed annotation guidelines
  - In-house developed annotation tool
- Workflow:
  1. Calibration round on 15 documents + discussion + consenzus
  2. Round 1
     - Three pairs of annotators, each working on 45 documents
     - Each annotator annotated the data independently
  3. Round 2
     - Same as Round 1, but with reshuffled annotator pairs
  4. Estimate of the average pairwise IAA $\Rightarrow$ **70% agreement**
  5. Resolving the disagreements (one person)

Data Annotation

- News articles corpus of 285 documents
- Six trained annotators
  - Detailed annotation guidelines
  - In-house developed annotation tool
- Workflow:
  1. Calibration round on 15 documents + discussion + consenzus
  2. Round 1
     - Three pairs of annotators, each working on 45 documents
     - Each annotator annotated the data independently
  3. Round 2
     - Same as Round 1, but with reshuffled annotator pairs
  4. Estimate of the average pairwise IAA ⇒ **70% agreement**
  5. Resolving the disagreements (one person)

⇒ Final dataset: **270 documents** with **13K CR relations**

Our Focus

1. We don't consider the mention detection but instead work on **gold mentions**
2. We consider only the IDENTITY relation, which accounts for 87% CR relations
3. IDENTITY is an equivalence relation, thus we want **clusters**

Constrained Mention-Pair Model

- A mention-pair model is a **binary classifier**
  - Predicts whether two given mentions refer to the same entity
- To produce clusters of coreferent mentions, we need to couple the mention-pair model with
  1. A heuristic for creating mention-pair instances
  2. A method for ensuring the transitivity of coreference relations (i.e., coherence of pairwise decisions)

Creating Mention-Pair Instances

- Considering all possible mention pairs is not feasible
  - **Too many instances**, the vast majority of which are **negative**
- We follow the approach by Ng and Cardie (2002) for creating training instances
  - **A positive instance** between a mention m_j and its closest preceding non-pronomial coreferent mention m_i
  - **Negative instances** by pairing m_j with all mentions in between m_j and its closest preceding coreferent mention m_i (i.e., with $m_{i+1}, \ldots, m_{j-1}$)

The Mention-Pair Model

- A non-linear SVM (RBF) with 16 binary/numerical features:

  1. **String-matching features** compare two mentions at the superficial string level
     - strings identical, mention containment, longest common subsequence length, edit (Levenshtein) distance
  2. **Overlap features** quantify the overlap in tokens
     - at least one matching word/lemma/stem between mentions, number of common content (N/A/V/R) lemmas
  3. **Grammatical features** aim to indicate the grammatical compatibility of the mentions
     - pronominal mentions, gender match, number match
  4. **Distance-based features** measure how close are the mentions
     - distance in number of sentences/tokens, same sentence, adjacent mentions, number of mentions in between

Enforcing Transitivity

- By making only pairwise predictions, the mention-pair model **does not guarantee** document-level coherence of coreference
- We employ **constrained optimization via integer linear programming (ILP)** to ensure that document-level coreference transitivity holds

Enforcing Transitivity

- By making only pairwise predictions, the mention-pair model **does not guarantee** document-level coherence of coreference
- We employ **constrained optimization via integer linear programming (ILP)** to ensure that document-level coreference transitivity holds
- Objective function (to be maximized):

$$\sum_{(m_i,m_j)\in P} x_{ij} \cdot r(m_i,m_j) \cdot C(m_i,m_j)$$

- $r(m_i,m_j) \in \{-1,1\}$ is the mention-pair classifier's decision for mentions m_i and m_j
- $C(m_i,m_j) \in [0.5,1]$ is the confidence of the binary mention-pair classifier
- $x_{ij} \in \{0,1\}$ is the final decision for mentions m_i and m_j

Enforcing Transitivity

- Transitivity property is encoded via linear constraints

$$\begin{aligned} x_{ij} + x_{jk} - x_{ik} &\leq 1, \\ x_{ij} + x_{ik} - x_{jk} &\leq 1, \\ x_{jk} + x_{ik} - x_{ij} &\leq 1, \\ \forall \{(m_i, m_j), (m_j, m_k), (m_i, m_k)\} &\subseteq P \end{aligned}$$

- After optimization, we obtain **coreference clusters** by simply computing the **transitive closure** over **coherent** pairwise decisions x_{ij}

Experimental Setup

- Dataset split: 220 training documents, 50 test documents
- SVM model selection (C and γ optimization) using 10-fold CV on the train set

Experimental Setup

- Dataset split: 220 training documents, 50 test documents
- SVM model selection (C and γ optimization) using 10-fold CV on the train set
- Two baselines:
  1. Overlap baseline classifies mentions as coreferent if they share at least one content word
  2. GendNum baseline links each mention to the closest preceding mention of matching gender and number
- Standard closest-first clustering (Soon et al., 2001) is applied for both baselines

Experimental Setup

- Standard coreference evaluation metrics: MUC and B^3
- Models:
  - MP–$\textsc{Morph}$ – the binary mention pair model without grammatical features
  - MP – the binary mention-pair model
  - $\mathrm{MP{+}ILP}$ – the constrained mention-pair model (global coherence)

Results

	MUC			B^3		
Model	P	R	F_1	P	R	F_1
OVERLAP	81.0	42.9	54.1	75.7	54.5	61.4
GENDNUM	55.2	39.0	45.4	59.8	50.5	54.3
MP–MORPH	90.6	61.1	72.1	86.2	67.3	74.6
MP	89.4	**64.7**	74.2	84.0	**70.1**	75.4
MP+ILP	**91.9**	63.5	**74.4**	**90.6**	68.7	**77.6**

Results

	MUC			B^3		
Model	P	R	F_1	P	R	F_1
Overlap	81.0	42.9	54.1	75.7	54.5	61.4
GendNum	55.2	39.0	45.4	59.8	50.5	54.3
MP–Morph	90.6	61.1	72.1	86.2	67.3	74.6
MP	89.4	**64.7**	74.2	84.0	**70.1**	75.4
MP+ILP	**91.9**	63.5	**74.4**	**90.6**	68.7	**77.6**

- MP significantly outperforms the baselines
- Removing morphological features lowers F1 by $\sim$2 points
- Enforcing transitivity significantly increases B^3 by $\sim$2 points

Error Analysis

- Most false negatives due to cases where **external knowledge is needed**
  - *željezni kancelar* (*iron chancellor*)
    $\Updownarrow$
    Bismarck
- Most false positives due to **non-coreferent mentions with significant lexical overlap**
  - *Društvo hrvatskih književnika* (*Croatian Writers' Association*)
    $\Updownarrow$
    svečanosti u Društvu hrvatskih književnika
    (*ceremonies at the Croatian Writers' Association*)

Conclusion

- The first coreference resolution model for Croatian
- A **supervised mention-pair model** is coupled with **constrained optimization (using ILP)** to enforce transitivity of coreference relations
- Most errors originate from **lack of external knowledge** needed to infer coreference

Future Work

- **Knowledge-based features** from external sources like Wikipedia
- Detection of **near-identity** coreference relations (e.g., meronymy and zero anaphora)
- Building an **end-to-end coreference resolution system**
- A model for automated **mention detection**

Thanks!

Test data:

http://takelab.fer.hr/data/crocoref

TakeLab

http://takelab.fer.hr