Coreference Resolution Revisited

Mihai Surdeanu

msurdeanu@email.arizona.edu

www.surdeanu.info/mihai

April 19th, 2013
What is it?

Coreference resolution = the task of clustering together of expressions that refer to the same entity/concept.

Michelle LaVaughn Robinson Obama is an American lawyer and writer. She is the wife of the 44th and current President of the United States, Barack Obama, and the first African-American First Lady of the United States.
Why is it important?

• Question answering
  • “Who is Barack Obama’s spouse?”

• Information extraction
  • “Find all per:spouse relations between all named entities in this large corpus.”

• News aggregation
  • “What are recent events involving Michelle Obama?”
  • Requires cross-document coreference resolution. More on this soon.
Why is it important?

- Performance doubles for these applications when coreference resolution is used.

A typical algorithm

Identify all mentions
A typical algorithm

Compute link scores between all pairs
A typical algorithm

Partition this graph into entity clusters
Some insights learned

• Most algorithms focus on step 2: computing mention-pair scores using machine learning, which is a *local* operation
  • Poor representation of context: only two mentions considered
• Recent work showed that it is important to address coreference resolution as a *global* task, where all mentions are modeled jointly
  • This is hard to model using machine learning
• ML models generalize poorly to new words, domains, and languages
  • Annotating coreference is expensive
Idea 1

❌ machine learning
✔ deterministic, rule-based model
✔ “baby steps” approach
✔ global model
A senior adviser to Mitt Romney lashed out at President Barack Obama's re-election campaign Friday, accusing the president of waging a campaign based on "ugly distortions and lies."

"They have gone from what started out as petty distortions and untruths to unbelievable exaggerations that diminish the office of the president and insult the American people," Romney senior adviser Eric Fehrnstrom told reporters at a media briefing at the campaign's headquarters.

Events are fundamental for cross-document news aggregation!
News aggregation is a big business
Entities help event resolution

Doc 1

The New Orleans Saints placed Reggie Bush on the injured list or Wednesday.

Doc 2

Saints put Bush on I.R.
Events help entity resolution

Doc 1

One of the key suspected Mafia bosses arrested yesterday has hanged himself.

Doc 2

Police said Lo Presti had hanged himself.

Coreferent events have coreferent arguments!
Idea II

✔ joint entity and event coreference model
Overview

- Idea I – rule-based, global entity coreference resolution model
- Idea II – joint entity/event coreference resolution model
Overview

• Idea I – rule-based, global entity coreference resolution model
• Idea II – joint entity/event coreference resolution model
Entity coreference resolution model

- Novel architecture for coreference resolution:
  - **“Baby steps”** – accurate things first
  - **Global** – attribute sharing in clusters
  - **Deterministic** – rule-based model
- Top ranked system at CoNLL-2011 Shared Task:
  - 58.3% (open), 57.8% (closed)
Multiple passes (or “sieves”) over text
- Precision of each pass is smaller than preceding ones
- Recall keeps increasing with each pass
- Decisions once made cannot be modified by later passes
- Modular architecture
Baby-steps approach

- Multiple passes (or “sieves”) over text
- Precision of each pass is smaller than preceding ones
- Recall keeps increasing with each pass
- Decisions once made cannot be modified by later passes
- Modular architecture

“shaping”
“cautious learning”
“scaffolding”
Focus on high recall

Recall increases (precision decreases) as more sieves added

More global decisions

Post processing
The second attack occurred after some rocket firings aimed, apparently, toward the Israelis, apparently in retaliation. We’re checking our facts on that one. ... the strike will undermine efforts by Palestinian authorities to bring an end to terrorist attacks and does not contribute to the security of Israel.
Why multiple sieves?

The second attack occurred after some rocket firings aimed, apparently, toward the Israelis, apparently in retaliation. We’re checking our facts on that one. ... the strike will undermine efforts by Palestinian authorities to bring an end to terrorist attacks and does not contribute to the security of Israel.
The second attack occurred after some rocket firings aimed, apparently, toward the Israelis, apparently in retaliation. We're checking our facts on that one. ... The strike will undermine efforts by Palestinian authorities to bring an end to terrorist attacks and does not contribute to the security of Israel.
The second attack occurred after some rocket firings aimed, apparently, toward the Israelis, apparently in retaliation. We’re checking our facts on that one. ... the strike will undermine efforts by Palestinian authorities to bring an end to terrorist attacks and does not contribute to the security of Israel.
Pass 1 – Mention detection

• Extract all noun phrases (NP) plus pronouns and named entities even in modifier position

• Remove non-referring expressions, e.g., generic “it”, with manually written patterns
  • E.g., It is possible that...
Pass 2 – Speaker identification

- Extract speakers and use the info for resolution
  - “…”, she said.

- Positive and negative constraints for following sieves:
  “I voted for Nader because he was most aligned with my values,” she said.
Exactly the same text:

…TWA's bid for USAir skeptically, seeing it as a ploy to pressure USAir into buying TWA.

The Shahab 3 ground-ground missile: the new addition to Iran’s military capabilities … developed the Shahab 3 ground-ground missile for defense purposes with capabilities ranging from …
Pass 4 – Relaxed string match

String match after dropping the text following the head word:

…Clinton… Clinton, whose term ends in January…
Pass 5 – Precise constructs

Appositives:

... but Bob Gerson, video editor of This Week in Consumer Electronics, says Sony conceives ...

Predicate nominatives:

Started three years ago, Chemical's interest-rate options group was a leading force in the field.

Role appositives:

... [[actress] Rebecca Schaeffer] ...
... [[painter] Pablo Picasso] ...
Pass 5 – Precise constructs

Relative pronouns:

… [the finance street [which] has already formed in the Waitan district] …

Acronyms:

Agence France Presse … AFP

Demonyms/Gentilics:

Israel… Israeli
Passes 6 – 9: Strict head match

The Japanese company already has 12% of the total camcorder market, ranking it third behind the RCA and Panasonic brands ... The company also plans to aggressively start marketing ... The electronics company...

- Coupled with various constraints:
  - No new information in mentions to be resolved
  - No location mismatch, “Lebanon” != “southern Lebanon”
  - No numeric mismatch, “people” != “around 200 people”
  - No i-within-i, e.g., [[Sony Corporation] of America]
Pass 10 – Relaxed head match

• Same constraints as above but anaphora head can match any word in the candidate cluster

“Sanders”
is compatible with the cluster:
{Sauls, the judge, Circuit Judge N. Sanders Sauls}
Pass 11 – Pronoun resolution

• Attributes must agree
  • Number
  • Gender
  • Person
  • Animacy

• Assigned using POS tags, NER labels, static list of assignments for pronouns

• Improved further using gender and animacy dictionaries of Bergsma and Lin (2006), and Ji and Lin (2009)
Post processing

- Discard singleton clusters
  - This is why we could maximize recall in mention detection!
- Discard shorter mentions in appositive patterns
- Discard mentions that appear later in copulative relations

- Implemented to comply with OntoNotes annotations
A run-through example

John is a musician. He played a new song. A girl was listening to the song. “It is my favorite,” John said to her.
John is a musician. He played a new song. A girl was listening to the song. “It is my favorite,” John said to her.
A run-through example

John is a musician. He played a new song. A girl was listening to the song. “It is my favorite,” John said to her.
A run-through example

John is a musician. He played a new song. A girl was listening to the song. “It is my favorite,” John said to her.
A run-through example

John is a musician. He played a new song. A girl was listening to the song. “It is my favorite,” John said to her.

Precise constructs
A run-through example

John is a musician. He played a new song. A girl was listening to the song. “It is my favorite,” John said to her.

Strict head match
A run-through example

John is a musician. He played a new song. A girl was listening to the song. “It is my favorite,” John said to her.

Pronoun resolution
John is a musician. He played a new song. A girl was listening to the song. “It is my favorite,” John said to her.
Mention selection in a given sieve

- In each sieve, we consider for resolution only mentions that are currently first in textual order in their cluster.
- Most informative!
Features are shared within clusters

- **Within a cluster:**
  - Union of all modifiers
  - Union of all head words
  - Union of all attributes: number, gender, animacy

- **Robustness to missing/incorrect attributes**

```plaintext
"a group of students"

number: singular

"five students"

number: plural

number: singular, plural
```
EXPERIMENTS
### Results on older corpora

<table>
<thead>
<tr>
<th>UNSUPERVISED</th>
<th>ACE 2004 Test</th>
<th>ACE NWIRE</th>
<th>MUC6</th>
</tr>
</thead>
<tbody>
<tr>
<td>This work</td>
<td>81</td>
<td>80.2</td>
<td>74.4</td>
</tr>
<tr>
<td>Haghighi and Klein (2009)</td>
<td>79.0</td>
<td>76.9</td>
<td>75.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SUPERVISED</th>
<th>Ace 2004 Test</th>
<th>ACE NWIRE</th>
<th>MUC6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Culotta et al. (2007)</td>
<td>79.3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bengston and Roth (2008)</td>
<td>80.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Finkel and Manning (2008) +G</td>
<td>-</td>
<td>74.5</td>
<td>64.3</td>
</tr>
</tbody>
</table>

*B³ F1 scores of different systems on standard corpora*
Results: CoNLL-2011 closed track

CoNLL score = (MUC F1 + B^3 F1 + CEAF F1) / 3
Results: CoNLL-2011 open track

CoNLL score = (MUC F1 + B3 F1 + CEAF F1) / 3
CoNLL-2012 shared task

• **Multilingual** unrestricted coreference resolution in OntoNotes
  • English, Chinese, Arabic
• Higher barrier of entry
  • 16 submissions vs. 23 submissions in 2011
• But there was significant progress
  • Best score for English increased from 58.3 to 63.4
CoNLL-2012 shared task

- Two out of the top three systems used our system
- Fernandes et al., PUC/IBM Brazil
  - Adapted our system to Chinese and Arabic
  - Reranked the output of our system
  - Best system overall
- Chen and Ng, UT Dallas
  - Adapted our system to Chinese and Arabic
  - Added two ML-based sieves to our system
  - Best for Chinese, top 3 overall
- Proof that our approach is multilingual
### Analysis: Importance of sharing features

<table>
<thead>
<tr>
<th>Model</th>
<th>CoNLL F1 in OntoNotes Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity-centric model</td>
<td>59.3</td>
</tr>
<tr>
<td>Mention-pair model</td>
<td>55.9</td>
</tr>
</tbody>
</table>
# Analysis: Importance of multiple sieves

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-pass model</td>
<td>59.3</td>
</tr>
<tr>
<td>Single-pass model</td>
<td>53</td>
</tr>
</tbody>
</table>

CoNLL F1 in OntoNotes Dev
## Analysis: Importance of features

<table>
<thead>
<tr>
<th>Feature</th>
<th>CoNLL F1</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>59.3</td>
<td></td>
</tr>
<tr>
<td>wo/ Number</td>
<td>56.7</td>
<td>-2.6</td>
</tr>
<tr>
<td>wo/ Gender</td>
<td>58.9</td>
<td>-0.4</td>
</tr>
<tr>
<td>wo/ Animacy</td>
<td>58.3</td>
<td>-1.0</td>
</tr>
<tr>
<td>wo/ NE</td>
<td>58.8</td>
<td>-0.5</td>
</tr>
</tbody>
</table>

CoNLL F1 in OntoNotes Dev
Error analysis

- Semantics, discourse
- Pronoun resolution
- Non-referential mentions
- Event mentions
- Misc
New pictures reveal the sheer power of that terrorist bomb... In **these photos obtained by NBC News**, the damage...
Under the laws of the land, the ANC remains an illegal organization, and its headquarters are still in Lusaka, Zambia.
Error analysis

When **you** become a federal judge, all of a sudden **you** are relegated to a paltry sum.
“Support the troops, not the regime” That’s a noble idea until you’re supporting the weight of an armoured vehicle on your chest.
Idea I: Conclusions

• Novel architecture for coreference resolution
  • “Baby steps”
  • Global
  • Deterministic

• State of the art results (in multiple languages)
  • Best at CoNLL-2011
  • Two of the top 3 systems at CoNLL-2012 used it
Overview

• Idea I – rule-based, global entity coreference resolution model
• Idea II – joint entity/event coreference resolution model
Advanced Micro Devices announced its intention to buy ATI Technologies for $5.4 billion on Monday.

AMD announced the largest acquisition on Monday, paying about $5.4 billion to acquire ATI Technologies.
Entities help event resolution

Doc 1

The New Orleans Saints placed Reggie Bush on the injured list or Wednesday.

Doc 2

Saints put Bush on I.R.
Events help entity resolution

**Doc 1**

One of the key suspected Mafia bosses arrested yesterday *has hanged* himself.

**Doc 2**

Police said *Lo Presti* *had hanged* himself.

Coreferent events have coreferent arguments!
Objectives

• Goal
  • Holistic approach to coreference resolution: entities and events should be resolved \textit{jointly} and \textit{transparently}
  • Cross document

• Research questions
  • Does entity coreference help event coreference and vice versa?
  • Which features/model are best for this task?
Previous work

• Mostly about entity coreference
  • Ponzetto and Strube, 2006; Haghighi and Klein, 2009; Stoyanov et al., 2009; Haghighi and Klein, 2010; Raghunathan et al., 2010; Rahman and Ng, 2011

• A few about event coreference
  • Humphreys et al., 1997; Bagga and Baldwin, 1999; Chen and Ji, 2009; Bejan and Harabagiu, 2010

• Almost none on joint entity and event coreference
  • He 2007 – medical domain, focused on five semantic categories
Architecture

- Document Clustering
  - Broad topic detection
- Mention Extraction
  - Reduce search space
- High Precision Entity Resolution
  - “Baby-steps” resolution
- Entity & Event
- Pronoun Resolution
  - High-recall sieve
Document clustering

- Document Clustering
- Mention Extraction
- High Precision Entity Resolution
- Entity & Event
- Pronoun Resolution
Document clustering

• Reduces search space
  • Subsequent steps only work within a given cluster

• It provides a word sense disambiguation based on corpus-wide topics
  • hit in earthquake reports vs. criminal reports

• Non-parametric clustering model (Surdeanu et al., 2005)
  • EM variant
  • Initial points and number of clusters chosen using geometric heuristics
Example of document clustering

**Topic 1**

**Doc 1**

AMD announced the largest acquisition on Monday, paying about $5.4 billion to acquire ATI Technologies.

**Doc 2**

Advanced Micro Devices announced its intention to buy ATI Technologies for $5.4 billion on Monday.

**Topic 2**

**Doc 3**

... at least 40 people were injured in the earthquakes ...

**Doc 4**

A series of powerful earthquakes ... injuring dozens and destroying ...
Mention extraction

Document Clustering

Mention Extraction

High Precision Entity Resolution

Entity & Event

Pronoun Resolution
Mention extraction

• Nominal: same as the previous system
• Verbal: all VB* - some auxiliary/copulative verbs (e.g., “have”, “be”, “seem”)

• Note: events do appear as nominal mentions!
  • Hard to distinguish between nominal entity and event mentions
  • Our system transparently handles entity and event mentions
High-precision entity resolution

- All sieves minus the pronoun resolution sieve from previous system
- Further reduces search space
Iterative entity/event resolution

- Document Clustering
- Mention Extraction
- High Precision Entity Resolution
- Entity & Event
- Pronoun Resolution
Iterative entity/event resolution

Gold clusters
Iterative entity/event resolution

After high-precision sieves
Iterative entity/event resolution

- Calculate pairwise scores and merge best:
  \[(e_1, e_2) = \arg \max_{e_1, e_2 \in \mathcal{E}} \text{score}(e_1, e_2, \Theta)\]
  \[\mathcal{E}' = \text{merge}(e_1, e_2, \mathcal{E}')\]
- Regenerate features in modified clusters
- Transparently merges nominal and verbal mentions
Iterative entity/event resolution

• Calculate pairwise scores and merge best:
  \[(e_1, e_2) = \arg \max_{e_1, e_2 \in \mathcal{E}'} \text{score}(e_1, e_2, \Theta)\]
  \[\mathcal{E}' = \text{merge}(e_1, e_2, \mathcal{E}')\]

• Regenerate features in modified clusters
• Transparently merges nominal and verbal mentions
Iterative entity/event resolution

- Calculate pairwise scores and merge best:
  
  \[(e_1, e_2) = \arg \max_{e_1, e_2 \in \mathcal{E}'} \text{score}(e_1, e_2, \Theta)\]
  
  \[\mathcal{E}' = \text{merge}(e_1, e_2, \mathcal{E}')\]

- Regenerate features in modified clusters
- Transparently merges nominal and verbal mentions
Iterative entity/event resolution

- Calculate pairwise scores and merge best:
  \[(e_1, e_2) = \arg \max_{e_1, e_2 \in \mathcal{E}'} \text{score}(e_1, e_2, \Theta)\]
  \[\mathcal{E}' = \text{merge}(e_1, e_2, \mathcal{E}')\]

- Regenerate features in modified clusters
- Transparently merges nominal and verbal mentions
Iterative entity/event resolution

• Calculate pairwise scores and merge best:
  
  \[
  (e_1, e_2) = \arg \max_{e_1, e_2 \in \mathcal{E}'} \text{score}(e_1, e_2, \Theta)
  \]
  
  \[
  \mathcal{E}' = \text{merge}(e_1, e_2, \mathcal{E}')
  \]

• Regenerate features in modified clusters
• Transparently merges nominal and verbal mentions

Repeat
Iterative entity/event resolution

- Calculate pairwise scores and merge best:
  \[(e_1, e_2) = \arg \max_{e_1, e_2 \in \mathcal{E}'} \text{score}(e_1, e_2, \Theta)\]
  \[\mathcal{E}' = \text{merge}(e_1, e_2, \mathcal{E}')\]

- Regenerate features in modified clusters
- Transparently merges nominal and verbal mentions

Repeat
Iterative entity/event resolution

- Calculate pairwise scores and merge best:
  \[(e_1, e_2) = \arg \max_{e_1, e_2 \in \mathcal{E}'} \text{score}(e_1, e_2, \Theta)\]
  \[\mathcal{E}' = \text{merge}(e_1, e_2, \mathcal{E}')\]
- Regenerate features in modified clusters
- Transparently merges nominal and verbal mentions

Repeat
Iterative entity/event resolution

- Calculate pairwise scores and merge best:
  \[(e_1, e_2) = \arg \max_{e_1, e_2 \in \mathcal{E}'} \text{score}(e_1, e_2, \Theta)\]
  \[\mathcal{E}' = \text{merge}(e_1, e_2, \mathcal{E}')\]

- Regenerate features in modified clusters
- Transparently merges nominal and verbal mentions

Repeat
Iterative entity/event resolution

- Calculate pairwise scores and merge best:
  \[(e_1, e_2) = \arg \max_{e_1, e_2 \in \mathcal{E}'} \text{score}(e_1, e_2, \Theta)\]
  \[\mathcal{E}' = \text{merge}(e_1, e_2, \mathcal{E}')\]
- Regenerate features in modified clusters
- Transparently merges nominal and verbal mentions

Repeat
Iterative entity/event resolution

- Calculate pairwise scores and merge best:
  
  $$(e_1, e_2) = \arg \max_{e_1, e_2 \in \mathcal{E}'} \text{score}(e_1, e_2, \Theta)$$

  $$\mathcal{E}' = \text{merge}(e_1, e_2, \mathcal{E}')$$

- Regenerate features in modified clusters
- Transparently merges nominal and verbal mentions

Repeat
Iterative entity/event resolution

- Calculate pairwise scores and merge best:
  \[(e_1, e_2) = \arg\max_{e_1,e_2 \in \mathcal{E}'} \text{score}(e_1, e_2, \Theta)\]
  \[\mathcal{E}' = \text{merge}(e_1, e_2, \mathcal{E}')\]
- Regenerate features in modified clusters
- Transparently merges nominal and verbal mentions

Repeat
Training the model
Driven by two observations

1. Need regression because almost no generated cluster is perfectly correct or perfectly incorrect

\[ q = \frac{\text{links}_{\text{correct}}}{\text{links}_{\text{correct}} + \text{links}_{\text{incorrect}}} \]
Training the model

\[ q = 1.0 \]
\[ q = 0.33 \]
Training the model
Driven by two observations

2. Need an online learning setup because a brute force training approach generates an exponential number of candidates

- Training after applying the high-precision sieves
- Repeatedly applies the current model over training to generate candidates
- Model retrained after each epoch
- L2-regularized linear regression, 10 epochs
Features (1/3)

• Cosine similarity of vectors of head words (for nominal mentions) or head lemmas (for verbal mentions)
  • \{Barack Obama, President Obama, US president\} ➔ \{Obama:2, president:1\}

• The percentage of newly-introduced mention links after the merge that are WordNet synonyms
  • E.g., 2/6 for the merge of \{hit, strike\} and \{strike, join, say\}
Features (2/3)

- Number of coreferent arguments/predicates
  - E.g., 2 for $\text{AMD}_{\text{Arg0}} \text{ bought } \text{ATI}_{\text{Arg1}}$ and $\text{AMD}_{\text{Arg0}} \text{ acquired } \text{ATI}_{\text{Arg1}}$
- Number of coreferent arguments/predicates with a specific role (Arg0, Arg1, etc.)
  - E.g., 1 for Arg0 in the previous example
Features (3/3)

• 2nd order distributional similarity of mention words
  • E.g., the singleton cluster {a new home} becomes: {new:1, original:1, old:1, existing:1, current:1, unique:1, modern:1, different:1, special:1, major:1, small:1, home:1, house:1, apartment:1, building:1, hotel:1, residence:1, office:1, mansion:1, school:1, restaurant:1, hospital:1}

• Cosine similarity of number, gender, animacy, and NE label vectors
  • E.g., the vector for the cluster {systems, a pen} is: {number:singular:1, number:plural:1, gender:neutral:2}
Pronoun resolution

- From CoNLL-2011
- Needed because the previous components focus on nominal and verbal mentions
EXPERIMENTS
Corpus

- 43 topics, 482 documents from (Bejan, 2010)
- Annotated by 4 experts
- 5447 entity mentions, 2533 event mentions
  - Event mentions extend Bejan’s annotations; corrections made to align them to Onto Notes spec
  - Entity mentions annotated from scratch in house
- Example:

  A publicist says Tara Reid has checked herself into rehab …

  The beautiful party girl Tara Reid is taking the time this season and checking herself into rehab.
Baselines

• Baseline 1 – wo/ SRL
  • CoNLL-2011 model (for nominal and pronominal)
  • Lemma matching (for verbal)

• Baseline 2 – with SRL
  • Baseline 1 + two sieves
  • Merges two nominal clusters if head words match and predicates have same lemma
    • \{Obama_{Arg0} attended, the president_{Arg1} was elected\},
    \{Obama_{Arg1} was elected\}
  • Merges two verbal clusters if same lemma and arguments with same head word
  • Shows how much simple usage of argument info helps
## Results – All mentions

<table>
<thead>
<tr>
<th>Baseline 1 wo/ SRL</th>
<th>49.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline 2 with SRL</td>
<td>52.6</td>
</tr>
<tr>
<td>Complete model</td>
<td>55.9</td>
</tr>
</tbody>
</table>

CoNLL F1 on the test partition
## Results – Only entity mentions

<table>
<thead>
<tr>
<th>Model Description</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline 1 wo/ SRL</td>
<td>47.9</td>
</tr>
<tr>
<td>Baseline 2 with SRL</td>
<td>50.8</td>
</tr>
<tr>
<td>Complete model</td>
<td>54.2</td>
</tr>
</tbody>
</table>

CoNLL F1 on the test partition
Results – Only event mentions

<table>
<thead>
<tr>
<th>Baseline 1 wo/ SRL</th>
<th>51.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline 2 with SRL</td>
<td>52.2</td>
</tr>
<tr>
<td>Complete model</td>
<td>54.8</td>
</tr>
</tbody>
</table>

CoNLL F1 on the test partition
Discussion

• Many SRL features get high weights (6 out of 10)

<table>
<thead>
<tr>
<th>Entity Feature</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity Heads – Proper</td>
<td>1.10</td>
</tr>
<tr>
<td>Coreferent Predicate for ArgM–LOC – Common</td>
<td>0.45</td>
</tr>
<tr>
<td>Entity Heads – Common</td>
<td>0.36</td>
</tr>
<tr>
<td>Coreferent Predicate for Arg0 – Proper</td>
<td>0.29</td>
</tr>
<tr>
<td>Coreferent Predicate for Arg2 – Common</td>
<td>0.28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Event Feature</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event Lemmas</td>
<td>0.45</td>
</tr>
<tr>
<td>Coreferent Argument for Arg1</td>
<td>0.19</td>
</tr>
<tr>
<td>Links between Synonym</td>
<td>0.16</td>
</tr>
<tr>
<td>Coreferent Argument for Arg2</td>
<td>0.13</td>
</tr>
<tr>
<td>Number of Coreferent Arguments</td>
<td>0.07</td>
</tr>
</tbody>
</table>
Error analysis

- Pronoun resolution: 36%
- Semantics beyond role frames: 20%
- Arguments of nominal events: 17%
- Cascaded errors: 7%
- Initial high-precision sieves: 6%
- Phrasal verbs: 6%
- Other: 8%
The forces killed [at least 40 people]... and [the dead] included ...
The attack on [the school] has caused widespread shock across Israel . . . while Israeli forces on Tuesday killed at least 40 people during an attack on [a United Nations-run school in Gaza].
Idea II: Conclusions

• Holistic model for cross-document coreference resolution
  • Jointly solves references to events and entities by handling both nominal and verbal mentions

• Model/features
  • Yet another “baby-steps” model
  • Events and entities linked through semantic role frames

• Joint modeling beneficial for both entities and events
Big-picture conclusions

• Understanding the problem is more important than machine learning
• Model things jointly when you can
Acknowledgements

• Joint work with Heeyoung Lee, Karthik Raghunathan, Marta Recasens, Nathanael Chambers, Yves Peirsman, Angel Chang, Sudarshan Rangarajan, and Dan Jurafsky.
THANK YOU! QUESTIONS?

“Matthews ... we’re getting another one of those strange ‘aw blah es span yol’ sounds.”