Joint Learning for Event Coreference Resolution

Jing Lu and Vincent Ng
Human Language Technology Research Institute
University of Texas at Dallas
Event Coreference Subtasks

Nelson Mandela has died at age 95.
The world has lost a great man, former Prime Minister Julia Gillard said.
Event Coreference Subtasks

• **Trigger Detection**: Determine trigger word and subtype

Life.Die

Nelson Mandela has **died** at age 95.
The world has **lost** a great man, **former** Prime Minister Julia Gillard **said**.

Life.Die  Personnel.End-Position  Contact.Broadcast
Event Coreference Subtasks

- **Anaphoricity Detection**: Determine if an event mention is anaphoric.

  Not Anaphoric

  Nelson Mandela has **died** at age 95.
  The world has **lost** a great man, **former** Prime Minister Julia Gillard **said**.

  Anaphoric       Not Anaphoric       Not Anaphoric
Event Coreference Subtasks

• **Coreference**: Determine which event mentions refer to the same real-world event

Nelson Mandela has **died** at age 95. The world has **lost** a great man, **former** Prime Minister Julia Gillard **said**.
These three event subtasks are **inter-dependent**
Trigger-Coreference Dependencies

Nelson Mandela has **died** aged 95 surrounded ....

The world has **lost** a great man...

**Life.Die**  **Personnel.End-Position**  **Life.Injury**  **NULL**
Trigger-Coreference Dependencies

Life.Die
Nelson Mandela has died aged 95 surrounded ....

The world has lost a great man...

Nelson Mandela has died aged 95 surrounded .... The world has lost a great man...

Coreferent event mentions must have same subtype
Trigger-Anaphoricity Dependencies

The world has lost a great man...

The world has lost a great man...


Anaphoric
The world has lost a great man...


Anaphoric

Anaphoric event mentions cannot have subtype NULL
The world has lost a great man, former... 

Non-anaphoric
Coref-Anaphoricity Dependencies

The world has lost a great man, former...

Non-anaphoric

Non-anaphoric event mentions cannot have antecedents
Top Event Coreference Resolvers in KBP’16

- Largely capture inter-dependencies with a pipeline architecture

- Weakness: error propagation

Lu and Ng (2016), Liu et al. (2016), Nguyen et al. (2016)
Goal

- Improve event coreference performance by jointly learning three subtasks: trigger detection, anaphoricity detection, and coreference
  - Motivated by Durrett and Klein’s (2014) model and its success in jointly modeling the key tasks in the entity stack
  - Event anaphoricity detection introduced as an auxiliary task to mediate the interactions between trigger detection and coreference
Plan for the Talk

• Models
  – Independent model (unary, per-task factors)
Plan for the Talk

• Models
  – **Independent model** (unary, per-task factors)
  – **Joint model** (higher-order, cross-task factors)
Plan for the Talk

• Models
  – **Independent model** (unary, per-task factors)
  – **Joint model** (higher-order, cross-task factors)

• Evaluation
Mandela has died... The world has lost... former Prime...
Mandela has died... The world has lost... former Prime Minister...
Mandela has died... The world has lost... former Prime Minister...
Mandela has died... The world has lost... former Prime Minister...
Model

- Log-linear probability distribution over the variables

\[ p(t, a, c | x; \Theta) \propto \exp\left( \sum_i \theta_i f_i(t, a, c, x) \right) \]
Model

- Log-linear probability distribution over the variables

\[ p(t, a, c|x; \Theta) \propto \exp\left( \sum_i \theta_i f_i(t, a, c, x) \right) \]
Model

• Log-linear probability distribution over the variables

\[ p(t, a, c | x; \Theta) \propto \exp \left( \sum_i \theta_i f_i(t, a, c, x) \right) \]
Model

• Log-linear probability distribution over the variables

\[ p(t, a, c | x; \Theta) \propto \exp(\sum_i \theta_i f_i(t, a, c, x)) \]

Input document context
Feature function
Weight vector

The world has lost...
Independent Model

\[ p(t, a, c|x; \Theta) \propto \exp(\Theta^T F(t, a, c, x)) \]

\[ F_{\text{Indep}} = \sum_{i=1}^{n} f(a_i, x) + f(c_i, x) + f(t_i, x) \]
The world has lost a great man, former Prime Minister...
The world has lost a great man, former Prime Minister...
The world has lost a great man, former Prime Minister...

- Trigger word: lost
- Trigger Lemma: lose
Trigger Detection Features

The world has lost a great man, former Prime Minister...

Trigger word: lost
Trigger Lemma: lose
Bigram: has_lost, lost_a
The world has lost a great man, former Prime Minister...

**Trigger Detection Features**

- **Trigger word**: lost
- **Trigger Lemma**: lose
- **Bigram**: has_lost, lost_a
- **Nearest Entity (Left)**: world
- **Nearest Entity (Right)**: man
Trigger Detection Features

The world has **lost** a great man, former Prime Minister...

| Trigger word | lost |
| Trigger Lemma | lose |
| Bigram | has_lost, lost_a |
| Nearest Entity (Left) | world |
| Nearest Entity (Right) | man |
| Dependencies | World, man |
The world has lost a great man, former Prime Minister...

| Trigger word | lost |
| Trigger Lemma | lose |
| Bigram | has_lost, lost_a |
| Nearest Entity (Left) | world |
| Nearest Entity (Right) | man |
| Dependencies | World, man |

Paired with trigger lemma
Event Coreference

The world has lost...

c2

\rightarrow died

NEW

Durrett and Klein 2013
Event Coreference Features

Features representing the current mention

The world has lost...

Durrett and Klein 2013
Event Coreference Features

Features representing the current mention

Features representing a preceding mention and its relationship with the current mention

The world has lost...

Durrett and Klein 2013
Anaphoricity Detection Features

The world has lost a great man, former Prime Minister...
Anaphoricity Detection Features

The world has lost a great man, former Prime Minister...

Trigger words of preceding mentions died
Anaphoricity Detection Features

The world has lost a great man, former Prime Minister...

Trigger words of preceding mentions: died
Same lemma: No
Anaphoricity Detection Features

The world has **lost** a great man, former Prime Minister...

- Trigger words of preceding mentions: **died**
- Same lemma: **No**
- Anaphoricity probability: **0.75**
Plan for the Talk

• Models
  – **Independent model** (unary, per-task factors)
  – **Joint model** (higher-order, cross-task factors)

• Evaluation
Joint Model

\[ p(t, a, c | x; \Theta) \propto \exp(\Theta^T F(t, a, c, x)) \]

\[ F_{\text{Indep}} = \sum_{i=1}^{n} f(a_i, x) + f(c_i, x) + f(t_i, x) \]

The world has lost...
Joint Model

\[ p(t, a, c|x; \Theta) \propto \exp(\Theta^T F(t, a, c, x)) \]

\[ F_{\text{Indep}} = \sum_{i=1}^{n} f(a_i, x) + f(c_i, x) + f(t_i, x) \]

\[ F_{\text{Joint}} = F_{\text{Indep}} + \sum \text{joint factors} \]

The world has lost...
Trigger – Coref Factors

Mandela has **died**... The world has **lost**...
Mandela has died... The world has lost...
Mandela has died... The world has lost...

Coreferent mentions should have the same subtype
The world has lost… Mandela has died… The world has lost…

Coreferent mentions should have the same subtype
Subtypes=Die-Die
The world has lost…

Mandela has died… The world has lost…

Coreferent mentions should have the same subtype

Subtypes=Die-Die
CurrType=Die, Prev = died
Mandela has died... The world has lost...
Trigger - Anaphoricity Factors

The world has lost...
The world has lost...
Trigger - Anaphoricity Factors

Anaphoric event mention cannot have subtype NULL

Subtypes=Die, Lemma=lose, Anaphoricity = Yes

The world has lost...
Coref - Anaphoricity Factors

The world has lost...
Coref - Anaphoricity Factors

Anaphoric event mentions should have antecedents

The world has lost...
Coref - Anaphoricity Factors

Anaphoric event mentions should have antecedents

Anaphoricity = Yes, antecedent = Yes

The world has lost...
Joint Model

Mandela has died... The world has lost...
Training

• Log-linear probability distribution over the variables

\[ p(t, a, c|x; \Theta) \propto \exp(\sum_i \theta_i f_i(t, a, c, x)) \]

• Conditional log-likelihood of the training data with L1 regularization

\[ L(\Theta) = \sum_{i=1}^{d} \log \sum_{\text{latent stuff}} p'(t_i^*, a_i^*, c_i^*|x_i; \Theta) + \lambda \|\Theta\|_1 \]
Training

- Log-linear probability distribution over the variables

\[ p(t, a, c|x; \theta) \propto \exp(\sum_i \theta_i f_i(t, a, c, x)) \]

- Conditional log-likelihood of the training data with L1 regularization

\[ L(\Theta) = \sum_{i=1}^{d} \log \left( \sum_{\text{latent stuff}} p'(t_i^*, a_i^*, c_i^*|x_i; \Theta) + \lambda \|\Theta\|_1 \right) \]

\[ p'(t, a, c|x_i; \theta) \propto p(t, a, c|x_i; \theta) \exp[\alpha_t l_t(t, t^*) + \alpha_a l_a(a, a^*) + \alpha_c l_c(c, C^*)] \]

Gimpel and Smith, 2010
Training

- Log-linear probability distribution over the variables

\[
p(t, a, c|x; \Theta) \propto \exp\left( \sum_i \theta_i f_i(t, a, c, x) \right)
\]

- Conditional log-likelihood of the training data with L1 regularization

\[
L(\Theta) = \sum_{i=1}^{d} \log \sum_{\text{latent stuff}} p'(t_i^*, a_i^*, c_i^*|x_i; \Theta) + \lambda \|\Theta\|_1
\]

\[
p'(t, a, c|x_i; \theta) \propto p(t, a, c|x_i; \theta) \exp[\alpha_t l_t(t, t^*) + \alpha_a l_a(a, a^*) + \alpha_c l_c(c, C^*)]
\]

Gimpel and Smith, 2010
Inference: Belief Propagation

• computes the marginals for a variable or a set of variables to which a factor connects
• encourages compatibility among component models
• To improve scalability
  – Restrict the domains of the coreference variables
  – Restrict the domains of the subtype variables
Evaluation

• Dataset:
  – KBP 2016 English and Chinese corpora

• Evaluation metrics:
  – **Event Coreference**: MUC, B₃, CEAFₑ, BLANC, and average of their F-scores
  – **Trigger detection**: F-score
  – **Anaphoricity detection**: F-score
Event Coreference Results

- **KBP2016 (Jing and Ng, 2016)**
  - English: 30.08
  - Chinese: 26.43

Comparison of AVG-F Score for **Indep** and **Joint** models.
Event Coreference Results

<table>
<thead>
<tr>
<th>Language</th>
<th>KBP2016 (Jing and Ng, 2016)</th>
<th>Indep</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>30.08</td>
<td>31.28</td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>26.43</td>
<td>25.84</td>
<td></td>
</tr>
</tbody>
</table>
Event Coreference Results

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>KBP2016 (Jing and Ng, 2016)</td>
<td>30.08</td>
<td>26.43</td>
</tr>
<tr>
<td>Indep</td>
<td>31.28</td>
<td>25.84</td>
</tr>
<tr>
<td>Joint</td>
<td>33.08</td>
<td>27.79</td>
</tr>
</tbody>
</table>
Trigger Detection Results

KBP2016 (Jing and Ng, 2016)

- **English**: F Score = 46.99
- **Chinese**: F Score = 40.01
Trigger Detection Results

- **KBP2016 (Jing and Ng, 2016)**
- **Indep**
- **Joint**

<table>
<thead>
<tr>
<th>Language</th>
<th>F Score</th>
<th>Indep</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>46.99</td>
<td>48.82</td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>40.01</td>
<td>39.82</td>
<td></td>
</tr>
</tbody>
</table>

7/31/2017
Trigger Detection Results

<table>
<thead>
<tr>
<th>Language</th>
<th>KBP2016 (Jing and Ng, 2016)</th>
<th>Indep</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>46.99</td>
<td>48.82</td>
<td>49.3</td>
</tr>
<tr>
<td>Chinese</td>
<td>40.01</td>
<td>39.82</td>
<td>40.53</td>
</tr>
</tbody>
</table>

F SCORE
Anaphoricity Detection Results

<table>
<thead>
<tr>
<th>Language</th>
<th>F Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>27.35</td>
</tr>
<tr>
<td>Chinese</td>
<td>19.31</td>
</tr>
</tbody>
</table>
Anaphoricity Detection Results

F SCORE

<table>
<thead>
<tr>
<th>Language</th>
<th>Indep</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>27.35</td>
<td>31.94</td>
</tr>
<tr>
<td>Chinese</td>
<td>19.31</td>
<td>23.33</td>
</tr>
</tbody>
</table>
Conclusion

• Joint predictions on trigger detection, event coreference resolution, and event anaphoricity

• Joint models effectively improve the performance of event coreference resolution
  – achieves the best results to date on KBP 2016 English and Chinese corpora
Model Ablation

• evaluate the importance of each of the three types of joint factors
  – add each type of joint factors to the independent model
  – remove each type of joint factors from the full joint model
# Model Ablation: Coref-Trigger interactions

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coref</td>
<td>Trigger</td>
</tr>
<tr>
<td><strong>INDEP.</strong></td>
<td>31.28</td>
<td>48.82</td>
</tr>
<tr>
<td><strong>INDEP.+CorefTrigger</strong></td>
<td>+0.39</td>
<td>+0.42</td>
</tr>
<tr>
<td><strong>INDEP.+CorefAnaph</strong></td>
<td>+0.40</td>
<td>−0.08</td>
</tr>
<tr>
<td><strong>INDEP.+TriggerAnaph</strong></td>
<td>+0.11</td>
<td>+0.38</td>
</tr>
<tr>
<td><strong>JOINT—CorefTrigger</strong></td>
<td>+0.56</td>
<td>−0.06</td>
</tr>
<tr>
<td><strong>JOINT—CorefAnaph</strong></td>
<td>+0.63</td>
<td>+0.66</td>
</tr>
<tr>
<td><strong>JOINT—TriggerAnaph</strong></td>
<td>+1.89</td>
<td>+0.50</td>
</tr>
<tr>
<td><strong>JOINT</strong></td>
<td>+1.80</td>
<td>+0.48</td>
</tr>
</tbody>
</table>

Contributes the most to coreference performance, and trigger detection
Model Ablation: Coref-Anaphoricity interactions

<table>
<thead>
<tr>
<th></th>
<th>Coref</th>
<th>Trigger</th>
<th>Anaph</th>
<th>Coref</th>
<th>Trigger</th>
<th>Anaph</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INDEP.</strong></td>
<td>31.28</td>
<td>48.82</td>
<td>27.35</td>
<td>25.84</td>
<td>39.82</td>
<td>19.31</td>
</tr>
<tr>
<td>INDEP.+CorefTrigger</td>
<td>+0.39</td>
<td>+0.42</td>
<td>−0.05</td>
<td>+0.95</td>
<td>+0.56</td>
<td>−0.37</td>
</tr>
<tr>
<td>INDEP.+CorefAnaph</td>
<td>+0.40</td>
<td>−0.08</td>
<td>+3.45</td>
<td>+0.37</td>
<td>+0.04</td>
<td>−0.11</td>
</tr>
<tr>
<td>INDEP.+TriggerAnaph</td>
<td>+0.11</td>
<td>+0.38</td>
<td>+1.35</td>
<td>+0.14</td>
<td>+0.52</td>
<td>+0.02</td>
</tr>
<tr>
<td><strong>JOINT—CorefTrigger</strong></td>
<td>+0.56</td>
<td>−0.06</td>
<td>+4.41</td>
<td>+0.19</td>
<td>+0.35</td>
<td>+3.34</td>
</tr>
<tr>
<td>JOINT—CorefAnaph</td>
<td>+0.63</td>
<td>+0.66</td>
<td>+1.46</td>
<td>+1.50</td>
<td>+0.88</td>
<td>+0.28</td>
</tr>
<tr>
<td><strong>JOINT—TriggerAnaph</strong></td>
<td>+1.89</td>
<td>+0.50</td>
<td>+4.01</td>
<td>+1.65</td>
<td>+0.50</td>
<td>+1.79</td>
</tr>
<tr>
<td><strong>JOINT</strong></td>
<td>+1.80</td>
<td>+0.48</td>
<td>+4.59</td>
<td>+1.95</td>
<td>+0.71</td>
<td>+4.02</td>
</tr>
</tbody>
</table>

In isolation: improve coreference performance
In combination: improves both tasks, particularly anaphoricity determination
Model Ablation: Trigger-Anaphorhicity interactions

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th></th>
<th></th>
<th>Chinese</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coref</td>
<td>Trigger</td>
<td>Anaph</td>
<td>Coref</td>
<td>Trigger</td>
<td>Anaph</td>
</tr>
<tr>
<td><strong>INDEPENDENT</strong></td>
<td>31.28</td>
<td>48.82</td>
<td>27.35</td>
<td>25.84</td>
<td>39.82</td>
<td>19.31</td>
</tr>
<tr>
<td>INDEP. + Coref Trigger</td>
<td>+0.39</td>
<td>+0.42</td>
<td>−0.05</td>
<td>+0.95</td>
<td>+0.56</td>
<td>−0.37</td>
</tr>
<tr>
<td>INDEP. + Coref Anaph</td>
<td>+0.40</td>
<td>−0.08</td>
<td>+3.45</td>
<td>+0.37</td>
<td>+0.04</td>
<td>−0.11</td>
</tr>
<tr>
<td>INDEP. + Trigger Anaph</td>
<td>+0.11</td>
<td>+0.38</td>
<td>+1.35</td>
<td>+0.14</td>
<td>+0.52</td>
<td>+0.02</td>
</tr>
<tr>
<td><strong>JOINT—Coref Trigger</strong></td>
<td>+0.56</td>
<td>−0.06</td>
<td>+4.41</td>
<td>+0.19</td>
<td>+0.35</td>
<td>+3.34</td>
</tr>
<tr>
<td>JOINT—Coref Anaph</td>
<td>+0.63</td>
<td>+0.66</td>
<td>+1.46</td>
<td>+1.50</td>
<td>+0.88</td>
<td>+0.28</td>
</tr>
<tr>
<td>JOINT—Trigger Anaph</td>
<td>+1.89</td>
<td>+0.50</td>
<td>+4.01</td>
<td>+1.65</td>
<td>+0.50</td>
<td>+1.79</td>
</tr>
<tr>
<td><strong>JOINT</strong></td>
<td>+1.80</td>
<td>+0.48</td>
<td>+4.59</td>
<td>+1.95</td>
<td>+0.71</td>
<td>+4.02</td>
</tr>
</tbody>
</table>

In isolation: improve both tasks
In combination: improves anaphoricity determination
Error Analysis

• Precision Error
  – Erroneous and mistyped triggers
    • assign the same subtype to event mentions triggered by the same word.
      – assigns the wrong subtype to triggers that possess different subtypes in different contexts
        » Example: trigger "leave" sometimes has the subtype Movement.Transport-Person and sometimes has the subtype Personnel.End-Position.
    • words that are only sometimes used as triggers are often wrongly posited as triggers when they are not
  – Failure to extract arguments
    • we only exploit dependency relations for argument extraction
    • Sometimes arguments that do not appear in the same sentence as their trigger.
Error Analysis

• Recall Error
  – Missing triggers
    • fails to identify trigger words that are unseen or rarely occurring in the training data.
  – Lack of entity coreference information.
    • the corresponding arguments of two coreferent event mentions are typically coreferent
    • This issue is particularly serious in discussion forum documents,
      – pronouns often serve as subjects and objects of event mentions
    • More serious in Chinese documents due to the zero pronouns
  – Lack of contextual understanding
    • Additional contextual information present in neighboring sentences is not used
      – when describing the fact that Tim Cook will attend Apple’s Istanbul store opening
      – “Cook is expected to return to Turkey for the store opening”
      – “Tim travels abroad YET AGAIN to be feted by the not-so-high-and-mighty”.

7/31/2017