Neural Networks and Coreference Resolution for Slot Filling

Heike Adel, Hinrich Schütze
Team CIS
University of Munich (LMU)

TAC workshop
November 16, 2015
CIS Slot Filling System: Overview

Improved Integration of Coreference Resolution

Relation Classification Models for Slot Filling

CIS Performance in the TAC Shared Task 2015
System overview

Query
(entity name + starting point)
System overview

Query
(entity name + starting point)

Alias component

Aliases for entity

Information retrieval component [Terrier]
System overview

Query
/entity name + starting point/

Alias component

Aliases for entity

Information retrieval component [Terrier]

Documents with aliases

Entity linking component [WAT]
System overview

Query
(entity name + starting point)

Alias component

Aliases for entity

Information retrieval component [Terrier]

Documents with aliases

Entity linking component [WAT]

Documents about entities

Candidate extraction component

Sentence extraction

Filler extraction

[Stanford CoreNLP]
System overview

Query
(entity name + starting point)

Alias component

Information retrieval component [Terrier]

Candidate extraction component

Candidate extraction component

Information retrieval component [Terrier]

Documents about entities

Documents with aliases

Entity linking component [WAT]

Sentence extraction

Filler extraction

[Stanford CoreNLP]

Possible slot fillers

Slot filler classification component

Aliases for entity

Alias component

Documents with aliases

Entity linking component [WAT]
System overview

Query (entity name + starting point)

Alias component

Aliases for entity

Documents about entities

Information retrieval component [Terrier]

Candidate extraction component

Documents with aliases

Entity linking component [WAT]

Sentence extraction [Stanford CoreNLP]

Filler extraction

Possible slot fillers

Slot filler classification component

Scored slot fillers

Postprocessing component

output
Contents of this talk

Query (entity name + starting point)

Alias component

Information retrieval component [Terrier]

Candidate extraction component [Stanford CoreNLP]

Slot filler classification component [WAT]

Entity linking component [WAT]

Documents with aliases

Documents about entities

Candidate extraction component

Sentence extraction [Stanford CoreNLP]

Filler extraction

Possible slot fillers

Scored slot fillers

Postprocessing component

output
How coreference could help slot filling

- Find every sentence with mentions of the entity
  ⇒ Provide models next in pipeline with all (?) necessary information to fill the slots
How coreference could help slot filling

- Find every sentence with mentions of the entity
  ⇒ Provide models next in pipeline with all (?) necessary information to fill the slots
- Get some slot fillers for free:
  - The mention “XX-year-old” already includes the fact that the entity is XX years old
    (same for “XX-based” or “XX-born”)
  - The mention “his mother” already includes the fact that the subject of the sentence is a child of the entity
How coreference could help slot filling

- Find every sentence with mentions of the entity
  ⇒ Provide models next in pipeline with all (?) necessary information to fill the slots
- Get some slot fillers for free:
  - The mention “XX-year-old” already includes the fact that the entity is XX years old
    (same for “XX-based” or “XX-born”)
  - The mention “his mother” already includes the fact that the subject of the sentence is a child of the entity

⇒ Coreference is a very important component of this task!
⇒ According to [Min and Grishman 2012, Pink et al. 2014], shortcomings of coreference resolution are one of the most important error sources!
Analysis: Shortcomings of coreference resolution systems

- Nominal anaphora like “XX-year-old”, “XX-based”, “XX-born” are not recognized as coreferent to the entity in the previous sentence in most cases.
Analysis: Shortcomings of coreference resolution systems

- Nominal anaphora like “XX-year-old”, “XX-based”, “XX-born” are not recognized as coreferent to the entity in the previous sentence in most cases
- Pronouns referring to the same entity are often clustered in the same chain - unfortunately, the entity is often clustered in another chain
  - Unlinked chains
  - Wrongly linked chains
Nominal anaphora: Improvements

▶ Heuristic:

```plaintext
Entity ∈ sentence_t
```

CIS at TAC: Neural Networks and Coreference Resolution for Slot Filling

Heike Adel

2015/11/16

7 / 21
Nominal anaphora: Improvements

- Heuristic:

  - Entity $\in$ $\text{sentence}_t$? 
    - yes
    - no

  - Nominal anaphor $\in$ $\text{sentence}_{t+1}$? 
    - yes
    - no

  - Ignore possible nominal anaphora
Nominal anaphora: Improvements

- Heuristic:

1. Entity ∈ sentence_t?
   - yes
   - no

2. Nominal anaphor ∈ sentence_{t+1}?
   - yes
   - no

3. Another entity directly after anaphor?
   - yes
   - no

Ignore possible nominal anaphora
Nominal anaphora: Improvements

- Heuristic:

![Flowchart diagram]

- Entity $\in$ sentence$_t$?
  - yes
  - Nominal anaphor $\in$ sentence$_{t+1}$?
    - yes
      - Another entity directly after anaphor?
        - yes
          - Nominal anaphor may refer to entity
        - no
          - Ignore possible nominal anaphora
    - no
      - yes
        - Ignore possible nominal anaphora
      - no
Expansion of coreference integration

- CIS SF system for 2014 evaluation: only coreference resolution for entities from queries (<name>)
- BUT: consider a sentence like “He is her father.”
Expansion of coreference integration

- CIS SF system for 2014 evaluation: only coreference resolution for entities from queries (<name>)
- BUT: consider a sentence like “He is her father.”
- Analysis: Coreference resolution for filler: important especially due to newly introduced inverse slots
  - 2014: 8 slots with PER fillers
  - 2015: 20 slots with PER fillers
- Future work: Investigate the effect of coreference resolution for fillers in more detail Extend it to other filler types as well
Expansion of coreference integration

- CIS SF system for 2014 evaluation: only coreference resolution for entities from queries (<name>)
- BUT: consider a sentence like “He is her father.”
- Analysis: Coreference resolution for filler: important especially due to newly introduced inverse slots
  - 2014: 8 slots with PER fillers
  - 2015: 20 slots with PER fillers
- Now: coreference resolution for both <name> and <filler>
  - But only if filler is a person
  - Future work: Investigate the effect of coreference resolution for fillers in more detail
    Extend it to other filler types as well
Coreference resource

- Observation: Long runtime of coreference resolution systems
- Solution: Corpus pre-processing
Coreference resource

- Observation: Long runtime of coreference resolution systems
- Solution: Corpus pre-processing
- TAC source corpus: \(~65\%\) pre-processed with [Stanford CoreNLP] so far
  - \(~30M\) chains and \(~105M\) mentions found
  - \(~25M\) pronoun mentions

Easily accessible format: chains of mention start offset - end offset pairs

NYT ENG

20090601.0015 14 2424-2441 87-95 170-178 812-820 890-892 1473-1483 1785-1793 2036-2044 2493-2495 211-250 1649-1657 798-892 587-595 1121-1129 1130-1132...

Resource will be publicly available
Coreference resource

- Observation: Long runtime of coreference resolution systems
- Solution: Corpus pre-processing
- TAC source corpus: \( \sim 65\% \) pre-processed with [Stanford CoreNLP] so far
  - \( \sim 30\)M chains and \( \sim 105\)M mentions found
  - \( \sim 25\)M pronoun mentions
- Easily accessible format: chains of mention start offset - end offset pairs
  - NYT\_ENG\_20090601.0015 14
    2424-2441 87-95 170-178 812-820 890-892 1473-1483
    1785-1793 2036-2044 2493-2495
    211-250 1649-1657
    798-892 587-595 1121-1129 1130-1132
    ...
- Resource will be publicly available
Classification component 2015

- **Data available?**
  - yes
  - no

  **Pattern matcher [Roth 2013]**
  - match: 1.0
  - no match: 0.0

  **SVM**
  - probability
  - weighted sum

  **CNN**
  - probability

  **RNN**
  - probability

  **Result**
Convolutional neural networks: Motivation

- Extract most relevant n-grams
  - Convolution: Create n-gram representations
  - Pooling: Find most relevant n-grams
  - ... independent of position in sentence
- Use n-gram based sentence representation for classification
- Wordvectors: implicit handling of synonyms
CNNs for slot filling

- Input: pre-trained word embeddings [word2vec]
- Context splitting
- Convolution and pooling for all contexts separately
- MLP (one hidden layer) and softmax for relation classification
Recurrent neural networks: Motivation

- Create global sentence representation
- ... using all available information
- Possibly more robust against insertions (than e.g. patterns)
- Possibly better with longer sentence lengths (than CNN)
RNNs for slot filling

Uni-directional RNN

```
+-----------------+
| h   V           |
| U   h   V       |
| w_1 h   V       |
| <> h   V        |
| w_3 h   V       |
```

Relation

```
+-----------------+
| h               |
| U   h           |
| w_5 h           |
| <> h            |
| w_6 h           |
```

Input: pre-trained word embeddings [word2vec]

Softmax for classification

(1) Uni-directional RNN

(2) Bi-directional RNN

(3) Multi-task bi-directional RNN

Predict type of next word (rel_argument_1, rel_argument_2, other)

Result of RNN component: score of the most confident RNN
Performance in the TAC shared task 2015
CIS runs

- All runs include coreference resolution
- All runs: automatically tuned slot-wise output thresholds
CIS runs

- All runs include coreference resolution
- All runs: automatically tuned slot-wise output thresholds
- Submission of five runs:
  - Base run: classification with patterns + SVM + CNN
  - Non-neural run: base run - CNN
  - RNN run: base run + RNN
  - EL run: base run + entity linking for document extraction
  - High precision run: base run with output thresholds += 0.2
CIS runs

- All runs include coreference resolution
- All runs: automatically tuned slot-wise output thresholds
- Submission of five runs:
  - Base run: classification with patterns + SVM + CNN
  - Non-neural run: base run - CNN
  - RNN run: base run + RNN
  - EL run: base run + entity linking for document extraction
  - High precision run: base run with output thresholds += 0.2
CIS runs

- All runs include coreference resolution
- All runs: automatically tuned slot-wise output thresholds
- Submission of five runs:
  - Base run: classification with patterns + SVM + CNN
  - Non-neural run: base run - CNN
  - RNN run: base run + RNN
CIS runs

- All runs include coreference resolution
- All runs: automatically tuned slot-wise output thresholds
- Submission of five runs:
  - Base run: classification with patterns + SVM + CNN
  - Non-neural run: base run - CNN
  - RNN run: base run + RNN
  - EL run: base run + entity linking for document extraction
All runs include coreference resolution

All runs: automatically tuned slot-wise output thresholds

Submission of five runs:
- Base run: classification with patterns + SVM + CNN
- Non-neural run: base run - CNN
- RNN run: base run + RNN
- EL run: base run + entity linking for document extraction
- High precision run: base run with output thresholds $\pm 0.2$
CIS system results

- Best run: PAT + SVM + CNN + RNN
- Final results:

<table>
<thead>
<tr>
<th>Run Type</th>
<th>Mean</th>
<th>Macro</th>
<th>Max Macro</th>
<th>Max Micro</th>
</tr>
</thead>
<tbody>
<tr>
<td>high P run</td>
<td>12.87</td>
<td>14.01</td>
<td></td>
<td>13.77</td>
</tr>
<tr>
<td>base run</td>
<td>20.15</td>
<td>21.89</td>
<td></td>
<td>19.70</td>
</tr>
<tr>
<td>RNN run</td>
<td><strong>20.79</strong></td>
<td><strong>22.45</strong></td>
<td></td>
<td><strong>20.90</strong></td>
</tr>
<tr>
<td>EL run</td>
<td>20.39</td>
<td>22.15</td>
<td></td>
<td>20.21</td>
</tr>
<tr>
<td>non-neural run</td>
<td>17.60</td>
<td>19.28</td>
<td></td>
<td>14.62</td>
</tr>
</tbody>
</table>
Analysis 1: Impact of coreference resolution

- All submitted runs included coreference resolution
Analysis 1: Impact of coreference resolution

- All submitted runs included coreference resolution
- Offline run without coreference resolution
- Evaluated using the official assessments and scoring scripts

<table>
<thead>
<tr>
<th></th>
<th>hop 0 base run</th>
<th>hop 0 - coref</th>
<th>hop 1 base run</th>
<th>hop 1 - coref</th>
<th>all base run</th>
<th>all - coref</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>31.83</td>
<td>29.70</td>
<td>11.63</td>
<td>10.50</td>
<td>24.02</td>
<td>22.58</td>
</tr>
<tr>
<td>R</td>
<td>23.97</td>
<td>20.82</td>
<td>7.21</td>
<td>5.66</td>
<td>16.70</td>
<td>14.25</td>
</tr>
<tr>
<td>F1</td>
<td>27.35</td>
<td>24.48</td>
<td>8.90</td>
<td>7.36</td>
<td>19.70</td>
<td>17.47</td>
</tr>
</tbody>
</table>

⇒ Large impact of coreference resolution on end-to-end performance
Analysis 1: Impact of coreference resolution

- All submitted runs included coreference resolution
- Offline run without coreference resolution
- Evaluated using the official assessments and scoring scripts
- Results (max micro):

<table>
<thead>
<tr>
<th>Hop</th>
<th>Run Type</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>base run</td>
<td>31.83</td>
<td>23.97</td>
<td>27.35</td>
</tr>
<tr>
<td>0</td>
<td>- coref</td>
<td>29.70</td>
<td>20.82</td>
<td>24.48</td>
</tr>
<tr>
<td>1</td>
<td>base run</td>
<td>11.63</td>
<td>7.21</td>
<td>8.90</td>
</tr>
<tr>
<td>1</td>
<td>- coref</td>
<td>10.50</td>
<td>5.66</td>
<td>7.36</td>
</tr>
<tr>
<td>all</td>
<td>base run</td>
<td>24.02</td>
<td>16.70</td>
<td>19.70</td>
</tr>
<tr>
<td>all</td>
<td>- coref</td>
<td>22.58</td>
<td>14.25</td>
<td>17.47</td>
</tr>
</tbody>
</table>

⇒ Large impact of coreference resolution on end-to-end performance
Analysis 2: Impact of neural networks

- Design of runs to immediately assess the impact of the neural networks

<table>
<thead>
<tr>
<th>Model Configuration</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>hop 0 PAT+SVM</td>
<td>18.99</td>
<td>22.32</td>
<td>20.52</td>
</tr>
<tr>
<td>hop 0 PAT+SVM+CNN</td>
<td>31.83</td>
<td>23.97</td>
<td>27.35</td>
</tr>
<tr>
<td>hop 0 PAT+SVM+CNN+RNN</td>
<td>29.98</td>
<td>26.58</td>
<td>28.18</td>
</tr>
<tr>
<td>hop 1 PAT+SVM</td>
<td>5.92</td>
<td>4.53</td>
<td>5.13</td>
</tr>
<tr>
<td>hop 1 PAT+SVM+CNN</td>
<td>11.63</td>
<td>7.21</td>
<td>8.90</td>
</tr>
<tr>
<td>hop 1 PAT+SVM+CNN+RNN</td>
<td>13.82</td>
<td>6.08</td>
<td>8.44</td>
</tr>
<tr>
<td>all PAT+SVM</td>
<td>14.64</td>
<td>14.60</td>
<td>14.62</td>
</tr>
<tr>
<td>all PAT+SVM+CNN</td>
<td>24.02</td>
<td>16.70</td>
<td>19.70</td>
</tr>
<tr>
<td>all PAT+SVM+CNN+RNN</td>
<td>25.53</td>
<td>17.69</td>
<td>20.90</td>
</tr>
</tbody>
</table>

⇒ Neural networks improve end-to-end performance with 6.28 F1 points
Analysis 2: Impact of neural networks

- Design of runs to immediately assess the impact of the neural networks

- Results (max micro):

<table>
<thead>
<tr>
<th>Hop</th>
<th>Model</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>hop 0</td>
<td>PAT+SVM</td>
<td>18.99</td>
<td>22.32</td>
<td>20.52</td>
</tr>
<tr>
<td>hop 0</td>
<td>PAT+SVM+CNN</td>
<td>31.83</td>
<td>23.97</td>
<td>27.35</td>
</tr>
<tr>
<td>hop 0</td>
<td>PAT+SVM+CNN+RNN</td>
<td>29.98</td>
<td>26.58</td>
<td>28.18</td>
</tr>
<tr>
<td>hop 1</td>
<td>PAT+SVM</td>
<td>5.92</td>
<td>4.53</td>
<td>5.13</td>
</tr>
<tr>
<td>hop 1</td>
<td>PAT+SVM+CNN</td>
<td>11.63</td>
<td>7.21</td>
<td>8.90</td>
</tr>
<tr>
<td>hop 1</td>
<td>PAT+SVM+CNN+RNN</td>
<td>13.82</td>
<td>6.08</td>
<td>8.44</td>
</tr>
<tr>
<td>all</td>
<td>PAT+SVM</td>
<td>14.64</td>
<td>14.60</td>
<td>14.62</td>
</tr>
<tr>
<td>all</td>
<td>PAT+SVM+CNN</td>
<td>24.02</td>
<td>16.70</td>
<td>19.70</td>
</tr>
<tr>
<td>all</td>
<td>PAT+SVM+CNN+RNN</td>
<td>25.53</td>
<td>17.69</td>
<td>20.90</td>
</tr>
</tbody>
</table>

⇒ Neural networks improve end-to-end performance with 6.28 F1 points
Conclusion

- Focus of this talk: coreference resolution, relation classification with neural networks
- Coreference resolution:
  - Coreference resolution for both relation arguments
  - Heuristical error post-processing
  \[ \Rightarrow \text{Considerable impact on end-to-end performance (esp. on recall)} \]
- Neural networks:
  - CNNs and RNNs
  - Interpolation of scores with non-neural model results
  \[ \Rightarrow \text{Very large impact on end-to-end performance} \]
Thanks for your attention!

Contact: heike.adel@cis.lmu.de
http://www.cis.uni-muenchen.de/~heike
References

- **Terrier:**

- **WAT:**

- **Stanford CoreNLP:**

- **Min and Grishman 2012:**
  Bonan Min, Ralph Grishman: Challenges in the knowledge base population slot filling task. In: LREC 2012.
References

- Pink et al. 2014:

- Roth 2013:

- word2vec:
Acknowledgements

- Heike Adel is a recipient of the Google Europe Fellowship in Natural Language Processing and this research is supported by this fellowship.
- This work was also supported by DFG (grant SCHU 2246/4-2).
- We would like to thank Pankaj Gupta for training the RNN models.