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

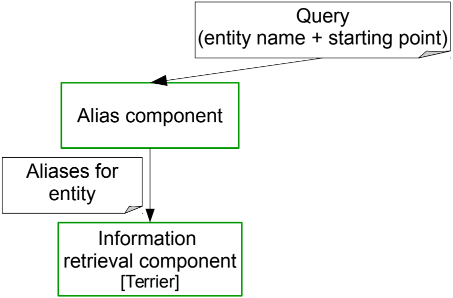
Relation Classification Models for Slot Filling

CIS Performance in the TAC Shared Task 2015

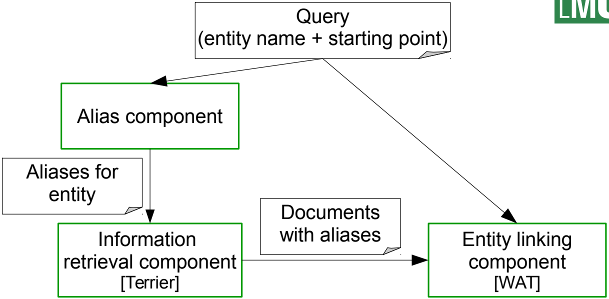
System overview

Query
(entity name + starting point)

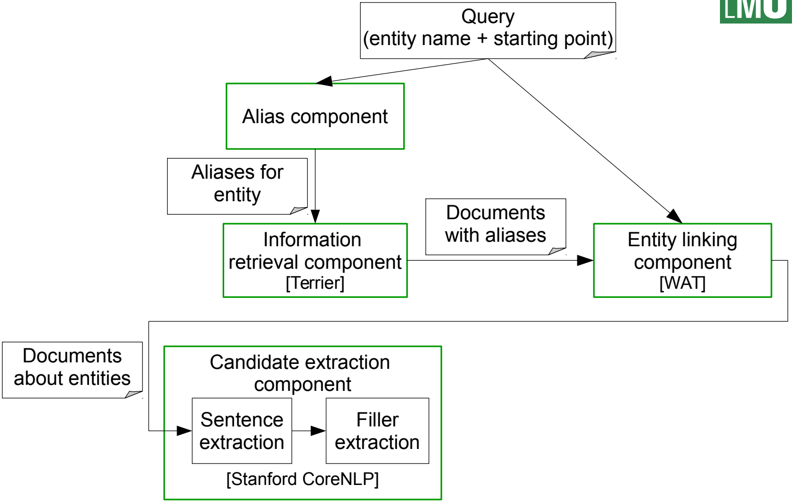
System overview



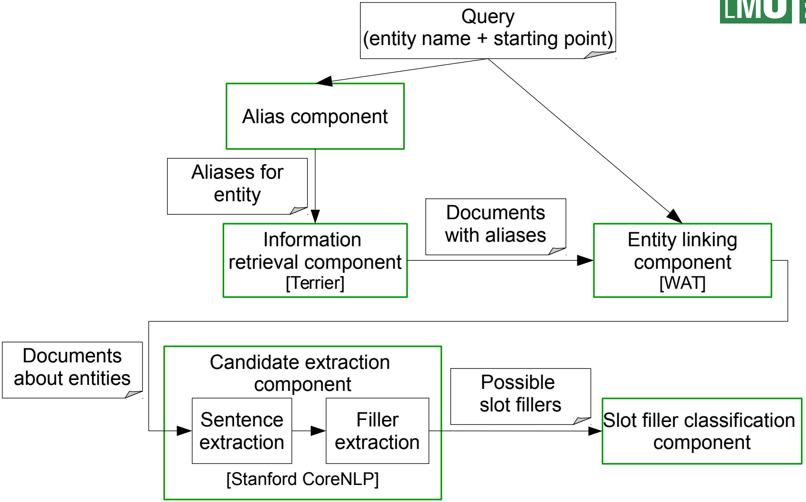
System overview



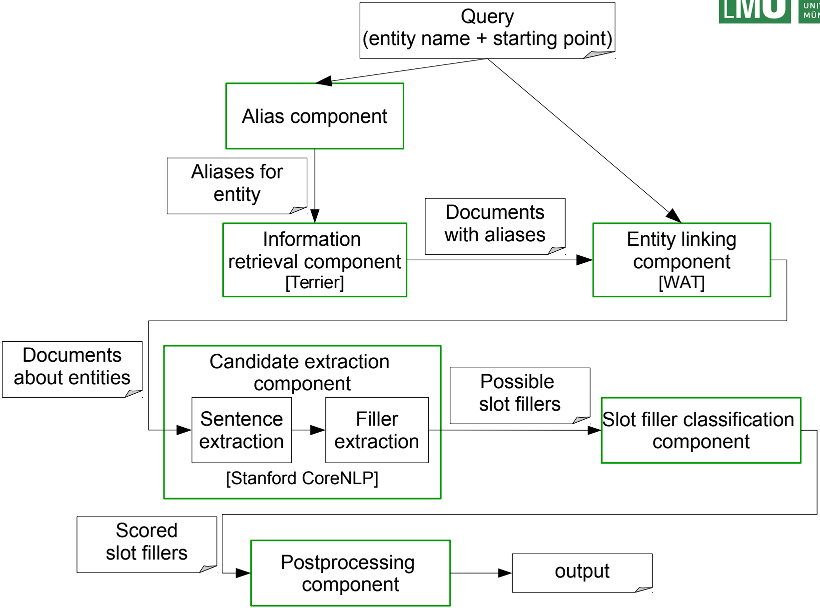
System overview



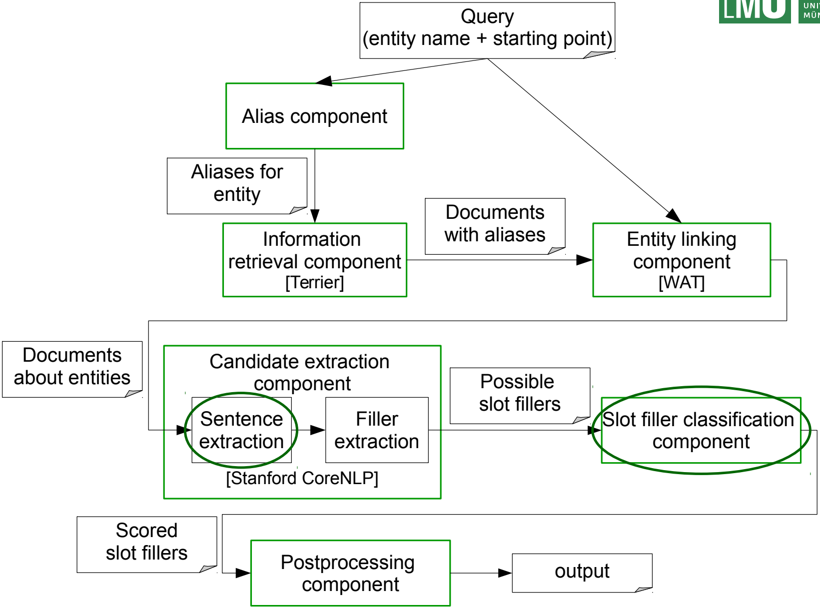
System overview



System overview



Contents of this talk



How coreference could help slot filling

- Find every sentence with mentions of the entity
  $\Rightarrow$ 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

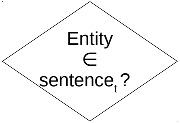
- Nominal anaphora like "XX-year-old", "XX-based", "XX-born" are not recognized as coreferent to the entity in the previous sentence in most cases

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

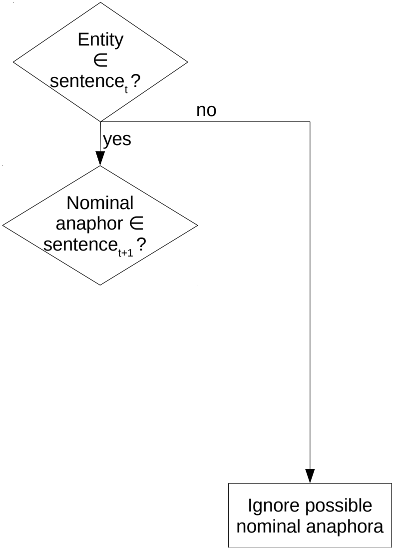
Nominal anaphora: Improvements

- Heuristic:



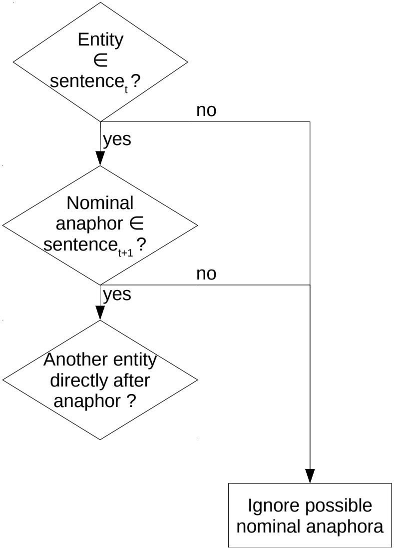
Nominal anaphora: Improvements

- Heuristic:



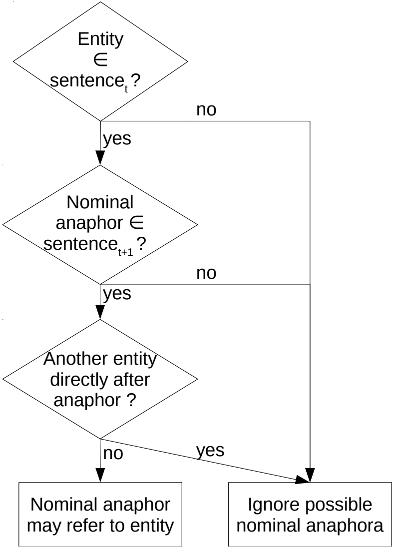
Nominal anaphora: Improvements

- Heuristic:



Nominal anaphora: Improvements

- Heuristic:



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

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

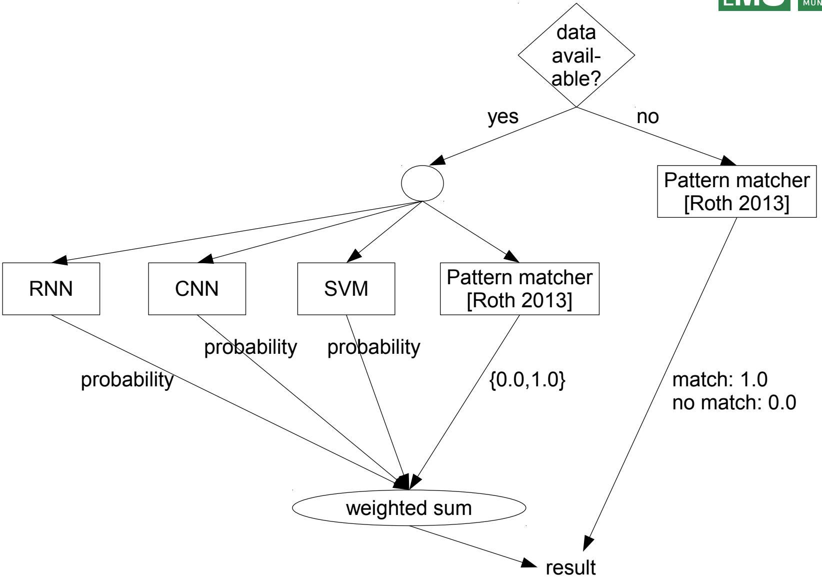
Coreference resource

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

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
    ```

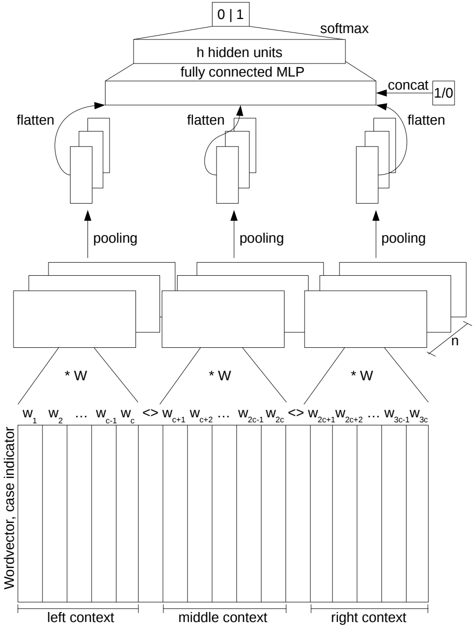
Classification component 2015



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



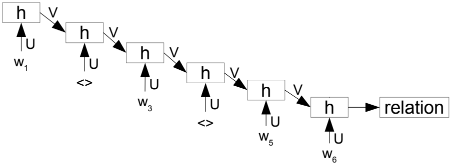
- Input: pre-trained word embeddings [word2vec]
- Context splitting
- Convolution and pooling for all contexts separately
- MLP (one hidden layer) and softmax for relation classification

Recurrent neural networks: Motivation

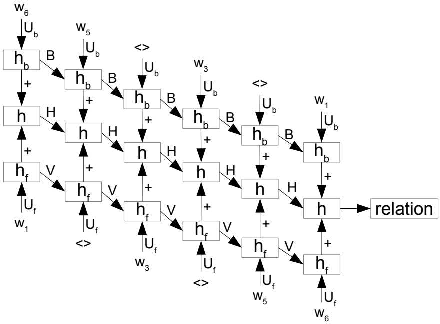
- Create global sentence representation
- ... using all available information
- Possibly more robust against insertions (than e.g. patterns)
- Possibly better with longer sentence lengths (than CNN)

RNNs for slot filling

Uni-directional RNN



Bi-directional RNN



- 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

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

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

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 system results

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

	mean macro	max macro	max micro
high P run	12.87	14.01	13.77
base run	20.15	21.89	19.70
RNN run	**20.79**	**22.45**	**20.90**
EL run	20.39	22.15	20.21
non-neural run	17.60	19.28	14.62

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

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):

		P	R	F1
hop 0	base run	**31.83**	**23.97**	**27.35**
hop 0	- coref	29.70	20.82	24.48
hop 1	base run	**11.63**	**7.21**	**8.90**
hop 1	- coref	10.50	5.66	7.36
all	base run	**24.02**	**16.70**	**19.70**
all	- coref	22.58	14.25	17.47

- $\Rightarrow$ 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

Analysis 2: Impact of neural networks

- Design of runs to immediately assess the impact of the neural networks
- Results (max micro):

		P	R	F1
hop 0	PAT+SVM	18.99	22.32	20.52
hop 0	PAT+SVM+CNN	**31.83**	23.97	27.35
hop 0	PAT+SVM+CNN+RNN	29.98	**26.58**	**28.18**
hop 1	PAT+SVM	5.92	4.53	5.13
hop 1	PAT+SVM+CNN	11.63	**7.21**	**8.90**
hop 1	PAT+SVM+CNN+RNN	**13.82**	6.08	8.44
all	PAT+SVM	14.64	14.60	14.62
all	PAT+SVM+CNN	24.02	16.70	19.70
all	PAT+SVM+CNN+RNN	**25.53**	**17.69**	**20.90**

- $\Rightarrow$ 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

  ⇒ Considerable impact on end-to-end performance (esp. on recall)
- Neural networks:
  - CNNs and RNNs
  - Interpolation of scores with non-neural model results

  ⇒ 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:
  Iadh Ounis, Gianni Amati, Vassilis Plachouras, Ben He, Craig Macdonald, Christina Lioma: Terrier: A high performance and scalable information retrieval platform. In: OSIR 2006.
- WAT:
  Francesco Piccinno, Paolo Ferragina: From Tagme to WAT: a new entity annotator. In: workshop on Entity recognition & disambiguation 2014.
- Stanford CoreNLP:
  Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, David McClosky: The Stanford CoreNLP natural language processing toolkit. In: ACL System Demonstrations 2014.
- 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:
  Glen Pink, Joel Nothman, James R Curran: Analysing recall loss in named entity slot filling. In: EMNLP 2014.
- Roth 2013:
  Benjamin Roth, Tassilo Barth, Michael Wiegand, Mittul Singh, Dietrich Klakow: Effective slot filling based on shallow distant supervision methods. In: TAC 2013.
- word2vec:
  Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean: Efficient estimation of word representations in vector space. In: Workshop at ICLR 2013.

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.