Singleton Detection using Word Embeddings and Neural Networks

Hessel Haagsma

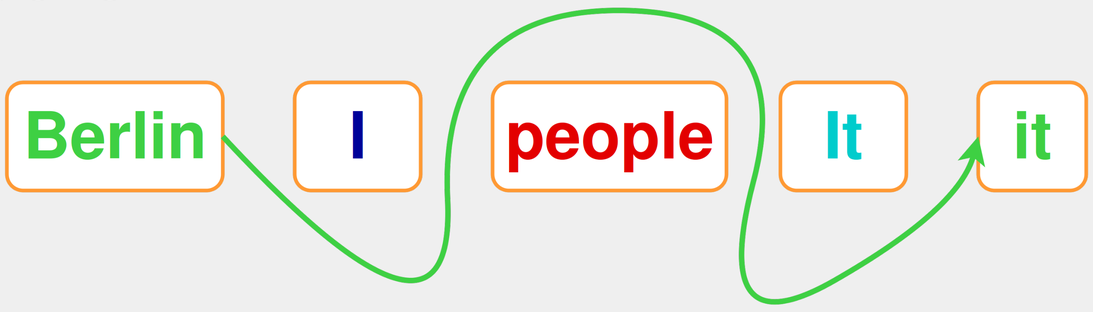
university of groningen

The Goal: Coreference Resolution

- **Example**

  $[\text{I}]_1$ have never been to $[\text{Berlin}]_2$ before.
  $[\text{It}]_3$ seems that $[\text{people}]_4$ really like $[\text{it}]_2$, though.

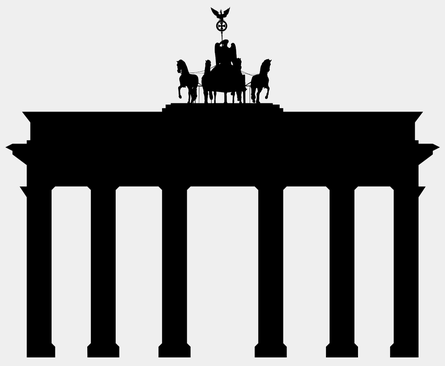
- **Singleton detection makes coreference resolution a lot easier by going from this**



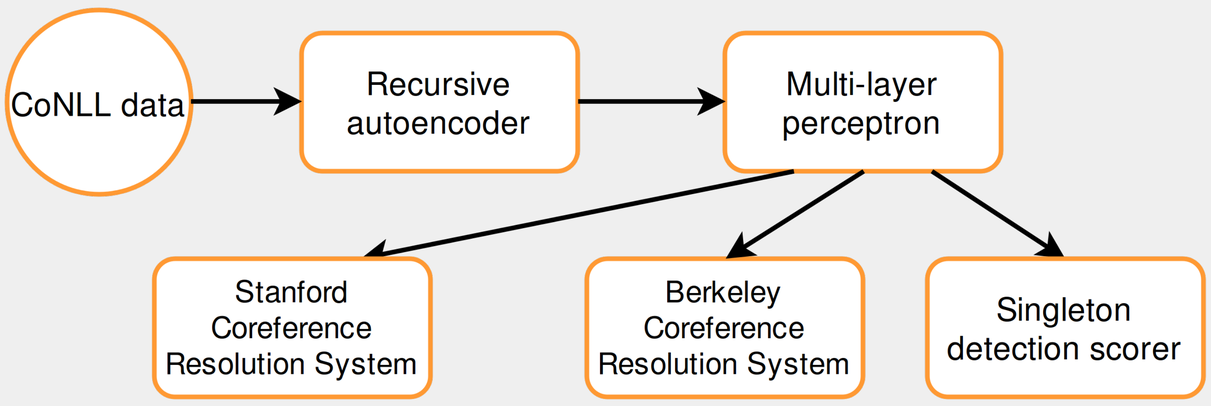
- **to this**

The Task: Singleton Detection

- **Singleton detection**
  Finding out which mentions do **not** participate in a coreference chain, *without* performing coreference resolution
- **Singletons**
  I, *It*, *people*
- **Coreferents**
  Berlin, *it*
- **Combines all previous mention-filtering tasks**
  non-referential *it* detection, non-anaphoric NP detection, non-antecedent NP detection, discourse-new mention detection



The Approach



- Word embeddings provide better generalization than lexicalized features
- Softmax classification allows for integration of probabilities into resolvers
- Mention representation by auto-encoding allows for easy integration of **neighbouring mentions as context**

The Data: Training and Testing

- **Existing coreference resolution datasets**
  Shared Tasks of CoNLL-2011 & CoNLL-2012
- **No singleton annotation**
  Use mentions as extracted by coreference resolution system
- **Data statistics**
  2011: 1.3M tokens - 332K mentions - 62.49% singleton
  2012: 1.6M tokens - 428K mentions - 56.22% singleton

Results: Singleton Detection

- **All results on CoNLL-2012 development set**

Model	Accuracy	Precision	Recall	F1
This work	**79.57**	79.77	85.83	**82.69**
De Marneffe et al.	79.0	**81.1**	80.8	80.9
Baseline	68.19	66.90	**87.21**	75.72

Results: Coreference Resolution

- **Stanford system results**

NE-filtering	Pair-wise	Threshold	CoNLL-F1
Yes	No	0.5	50.41*
Yes	No	0.15	56.73
Yes	Yes	0.5	55.46*
Yes	Yes	0.15	**57.17***
No	No	0.5	53.64*
No	No	0.15	56.71
No	Yes	0.5	55.96*
No	Yes	0.15	56.92*
Recasens et al.			56.90*
No Singleton Detection			56.44

- **Berkeley system results**

Model	CoNLL-F1
Probability	61.83
Probability + Binary (mentions)	61.81
Probability + Binary (pairs)	62.02
Probability + Binary (both)	62.02
De Marneffe et al.	**63.42**
No Singleton Detection	61.71

Interesting Examples

- **Good: captures patterns**
  And if you'd like to join me, **I***'d love it,*
  When the people went down and lifted **him** *up, he was dead.*
- **Bad: unknown words**
  Two ministers were so heavily implicated in the **Koskotas** *affair that [...]*
  Since NBC's interest in **the Qintex bid for MGM/UA** *was disclosed, [...]*

Conclusions

- **Singleton detection**
  Higher-dimensional embeddings improve performance (see paper)
  Mentions as context help, but only the closest ones (see paper)
  Unknown words are problematic, even with a 1.9M word vocabulary
- **Coreference resolution**
  High-precision filtering is the most effective way of integration
  Extrinsic evaluation is crucial, because **not all singletons are equal!**

Possible Future Improvements

- Separate classifiers for different mention types
- Use RNN, CNN, or LSTM, rather than MLP
- Improve treatment of unknown words, e.g. with POS-tags
- New state-of-the-art for singleton detection: 88.30% F1
  using SVM with surface features (Moosavi & Strube, NAACL2016)

hessel.haagsma@rug.nl