

---

# Supplementary Material: Ask Me Anything: Dynamic Memory Networks for Natural Language Processing

---

Ankit Kumar  
Ozan Irsoy  
Peter Ondruska  
Mohit Iyyer  
James Bradbury  
Ishaan Gulrajani  
Victor Zhong  
Romain Paulus  
Richard Socher

{james.bradbury,rpaulus,rsocher}@salesforce.com, Salesforce Inc., CA USA

## Mention Detection & Coreference Resolution

Coreference resolution is a complex structure prediction task which is usually solved in two steps of first detecting all mentions and then determining their coreference. Similarly to POS tagging we model mention detection as a sequence tagging problem but where the task is to predict a pair of symbols for every position corresponding to the number of mentions which start and end at given word (see Figure 1.). Next, the coreference for every pair of mentions is determined. Here, in turn we present DMN input module a text with highlighted single mention which coreference is classified with the mention provided as a part of the question. This formulation allows to ask the same question for different inputs and vice versa significantly speeding up evaluation of all  $N^2$  mention pairs.

We evaluated our coreference system using quizbowl dataset (Guha et al.) which showed to be difficult to solve using present state-of-the-art systems as it contains many nested mentions and complex entity relationships. The described mention detection performs here with precision 83.02, recall 75.12 and F1 score of 78.87 whereas the pairwise coreference classification of gold mentions achieves 68.35 precision, 63.90 recall and 66.05 F1 score. Moreover, as shown in Table 1, mention clusters obtained by the transitive closure of the predicted pairwise coreference graph outperform both (Guha et al.) and (Durrett & Klein, 2013) when evaluated using standard Conll metrics. Note that unlike these systems our approach does not use any hand-engineered features or text pre-processing such as to extract named entity tags or parse trees and to the best of our knowledge it is the first system which solves both mention detection and coreference resolution using purely deep learning methods.

Model	Metrics	$MUC$	$B^3$	$CEAF_e$
Guha et al.	P	56.8	68.1	73.3
	R	57.8	74.8	76.1
	F1	57.3	70.4	74.2
Durrett & Klein	P	70.2	88.5	56.5
	R	40.2	64.7	80.0
	F1	49.6	74.2	65.7
DMN	P	74.6	82.0	70.7
	R	66.0	77.5	79.3
	F1	<b>70.0</b>	<b>79.7</b>	<b>74.7</b>

Table 1. Coreference resolution scores of gold mentions from dataset of Guha et al., 2015 evaluated using Conll scorer v7.0.

## References

- Durrett, Greg and Klein, Dan. Easy victories and uphill battles in coreference resolution. In *EMNLP*, 2013.
- Guha, Anupam, Iyyer, Mohit, Bouman, Danny, Boyd-Graber, Jordan, and Boyd, Jordan. Removing the training wheels: A coreference dataset that entertains humans and challenges computers. In *North American Association for Computational Linguistics*.