COGNITIVELY AWARE COREFERENCE

Kellie Webster

Supervisor: James R. Curran

ə-lab, School of Information Technologies

kellie.webster@sydney.edu.au
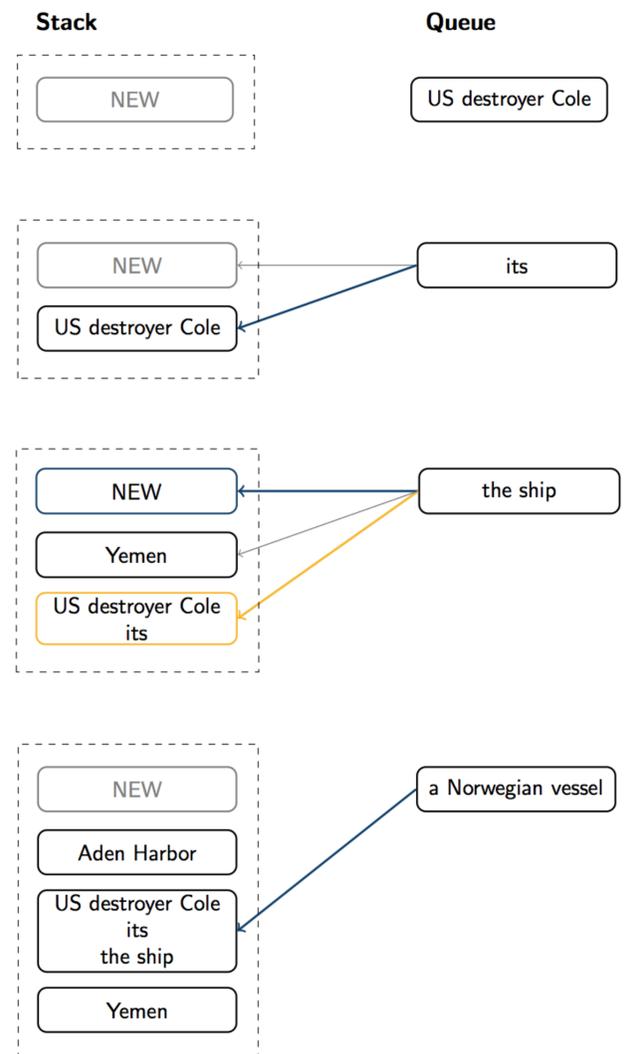
Coreference resolution

Authors refer to entities in the real world using noun mentions like **Cole** and **the ship**. Resolving the coreference between such mentions is a vital preprocessing step for smart information extraction systems to answer questions like:

- What ship is being towed out of **Aden Harbor**?

As human readers, we find this task easy, but it is very difficult to turn our ability into an automatic process.

- How do humans know that **Cole** and **the ship** refer to the same entity?
- Can we use facts about human cognitive processing to improve coreference resolution?

Current approaches

Coreference resolution systems typically train classifiers to label mention pairs as ± coreferential. The feature sets of such systems are good for proper names (*Cole*) and pronouns (*it*), but it is not clear how to design features for the more difficult case of nominals (*the ship*).

- How much do we rely on our **prior knowledge** of the Cole incident when reading this article?
- How much do we rely on **discourse patterns** to infer coreference?
- Does an author's style change to reflect how well known her topic is to her readership?

A further shortcoming of these mention-pair models is that they, by design, use only local cues in text as evidence of coreference. But:

- An article is a coherent discourse with information introduced in such a way that humans can understand its content.
- After reading this introduction, we know about Cole and have this information at hand as we read the remainder of the article.

LIMERIC

We introduced LIMERIC, our Limited Memory Incremental Coreference Resolution system, at the 25th International Conference on Computational Lingusitics this year [Webster and Curran, 2014].

We designed and built this system to capitalise on two key properties of cognitive processing:

Model simplicity

- Documents are processed in **human reading order** using a queue of extracted mentions; this enables us to model *information structure*.
- Clusters are stored in a **self-ordering** data structure in which order reflects the *cognitive accessibility* of a cluster's entity to the reader.

Model expressiveness

- Clusters of coreferential mentions are grown **incrementally**; attributes are shared to reflect global knowledge.
- Having access to this ordered collection enables us to devise **new features** expressing the relationship between a mention and these accessible entities.

Evaluation

	CoNLL F_1
Durrett and Klein [2013]	57.04
Chang et al. [2013]	61.02
Webster and Curran [2014]	**62.43**

Table 1: *Performance on OntoNotes 5*

We use the standard evaluation for coreference resolution using the OntoNotes 5 [Pradhan et al., 2007] dataset. OntoNotes 5 is a multi-lingual, multi-genre collection of over one million words of annotated data (English). This data comes from newswire, broadcast news and conversation, telephone conversation, and web text.

We benchmark against the best existing publicly available [Durrett and Klein, 2013] and research systems [Chang et al., 2013].

Processing a document

Conclusion

- LIMERIC represents a **new state-of-the-art** for coreference resolution.
- Cognitive insights both simplify our architecture and enable us to learn more expressive models.
- We substantially reduce the memory requirement for coreference resolution; this makes it more attractive for application downstream.

References

K. Chang, R. Samdani, and D. Roth. A constrained latent variable model for coreference resolution. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013.

G. Durrett and D. Klein. Easy victories and uphill battles in coreference resolution. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2013.

S. S. Pradhan, E. H. Hovy, M. P. Marcus, M. Palmer, L. A. Ramshaw, and R. M. Weischedel. Ontonotes: a unified relational semantic representation. *International Journal of Semantic Computing*, 1(4), 2007.

K. Webster and J. R. Curran. Limited memory incremental coreference resolution. In *Proceedings of the 25th International Conference on Computational Linguistics*, 2014.