Term Project Proposal
CS 710 - Artificial Intelligence

Coreference Resolution in Biomedical Text:
Improving Performance with Ontologies and Domain Specific NER

Mason Edmison
University of Wisconsin-Milwaukee
10/25/2019

1 Introduction

2 Dataset and Evaluation

Due to limited resources and time, I will not be able to train the coreference systems used and will use the pre-trained models[1]. For evaluating the coref system, I will use the training dataset from the task Protein Coreference at BioNLP 2011. I will evaluate the system on both mention detection as well as mention linking to produce coreference links using the harmonic mean *F1* as the determining metric.

3 Performance

	Prec.	Rec.	$F1$
SpaCy Web Md	15.15%	8.37%	10.72%

---

[1]Both systems, *deep coref* and *neural coref* ship pre-trained on the CoNLL 2012 dataset

References

[1] Mercedes Arguello Casteleiro, George Demetriou, Warren Read, Maria Jesus Fernandez Prieto, Nava Maroto, Diego Maseda Fernandez, Goran Nenadic, Julie Klein, John Keane, and Robert Stevens. Deep learning meets ontologies: experiments to anchor the cardiovascular disease ontology in the biomedical literature. *Biomedical Semantics*, 9(13), 2018.

[2] Miji Choi, Karin Verspoor, and Justin Zobel. Analysis of coreference relations in the biomedical literature. In *Proceedings of the Australasian Language Technology Association Workshop 2014*, pages 134–138, Melbourne, Australia, November 2014.

[3] Kevin Clark and Christopher D. Manning. Improving coreference resolution by learning entity-level distributed representations. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 643–653, Berlin, Germany, August 2016. Association for Computational Linguistics.

[4] Mohammad Taher Pilehvar and Nigel Collier. Improved semantic representation for domain-specific entities. In *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*, pages 12–16, Berlin, Germany, August 2016. Association for Computational Linguistics.

[5] Roman Prokofyev, Alberto Tonon, Michael Luggen, Loic Vouilloz, Djellel Eddine Difallah, and Philippe Cudré-Mauroux. Sanaphor: Ontology-based coreference resolution. In *Proceedings of the 14th International Conference on The Semantic Web - ISWC 2015 - Volume 9366*, pages 458–473, Berlin, Heidelberg, 2015. Springer-Verlag.