What Coreference Chains Tell about Experimental Groups in (Pre-)Clinical Trials

Annika Schwitteck
Bielefeld University
aschwitteck@techfak.uni-bielefeld.de

Hendrik ter Horst
Bielefeld University
hterhors@techfak.uni-bielefeld.de

Matthias Hartung
Bielefeld University
mhartung@cit-ec.uni-bielefeld.de

Tag
Datum
Zeit
Raum

NLP techniques are increasingly used for literature-based knowledge discovery from scientific publications in technical domains such as biomedicine (Henry & McInnes 2017). In the BMBF-funded project PSINK[1], we aim at automatically processing publications reporting the outcomes of pre-clinical experiments in the spinal cord injury domain in order to extract key parameters that are mandatory for assessing the translation potential of pre-clinical trials into clinical therapy concepts. As the main outcomes of pre-clinical trials are often derived from comparisons of treatments or interventions in different experimental groups (Brazda et al. 2017), an important sub-problem concerns the recognition of the experimental groups involved. We argue that, in order to capture them both at mention and reference level, the task should be framed as a coreference resolution problem.

In this work, we evaluate the potential of using state-of-the-art coreference resolution (Clark & Manning 2015) off the shelf for experimental group detection. We analyze whether the coreference chains produced can be reliably linked to ground-truth experimental groups in an annotated clinical trials corpus (Summerscales et al. 2011). Our evaluation addresses aspects of correctness and completeness of predicted experimental groups at mention and reference level (the latter being largely neglected in related previous work (Ferracane et al. 2016)).

References: ● Henry, S. and McInnes, B.T. (2017): Literature Based Discovery: Models, Methods, and Trends. *J Biomed Inform* 74, 20–32. ● Brazda, N., et al. (2017): SCIO. An Ontology to Support the Formalization of Pre-Clinical Spinal Cord Injury Experiments. *Proc. of JOWO – Ontologies and Data in the Life Sciences.* ● Clark, C. and Manning, C.D. (2015): Entity-Centric Coreference Resolution with Model Stacking. *Proc. of ACL*: 1405–1415. ● Summerscales, R.L., Argamon, S., Bai, S., Hupert, J. and Schwartz, A. (2011): Automatic summarization of results from clinical trials. *Proc. of BIBM*: 372–377. ● Ferracane, E., et al. (2016): Leveraging coreference to identify arms in medical abstracts. An experimental study. *Proc. of LOUHI*: 86–95.

---

[1] http://www.psink.de

1