Tree-LSTMs for Scientific Relation Classification
Sean MacAvaney, Luca Soldaini, Arman Cohan, and Nazli Goharian
Information Retrieval Lab, Computer Science Department, Georgetown University
(firstname)@ir.cs.georgetown.edu

Problem Statement

- **Task**: Extract semantic relations from scientific literature
- **Relations**: Model-Feature, Part-Whole, Compare, Result, and Topic

The CCLINC Korean-to-English translation system consists of two core modules.

- **Data**: Manually-annotated abstracts from the ACL Anthology Reference Corpus
- **Results**: Subtask 1.1: 9\textsuperscript{th} (of 28), Subtask 1.2: 5\textsuperscript{th} (of 20)
- **Code available at**: https://github.com/Georgetown-IR-Lab/semeval2018-task7

Method

- **Approach**: Feed dependency paths between entities into Tree-LSTM
- **Parsing**: Stanford dependencies (spaCy)
- **Child-sum variant tree LSTM**
- Only syntactic root of entity
- **Features**:
  - Wiki + arXiv embeddings
  - Dependency labels
  - Part-of-speech tags
  - Entity length (where applicable)
  - Height in tree

Results & Conclusions

- **Height feature causes model to overfit to training data**
- **Combination of general-language embeddings and domain-specific helps**
- **Often confuses Usage Model-Feature, and Part-Whole**
- **Difficulty identifying Topic**