

Yale University  
Computer Science Talk

# Towards Efficient and Generalizable Natural Language Processing

Host: Dragomir Radev



Hao Peng

Tuesday – March 29, 2022  
4:00 p.m.

Zoom Presentation

**Abstract:** Large-scale deep learning models have become the foundation of today’s natural language processing (NLP). Despite their recent, tremendous success, they struggle with generalization in real-world settings, like their predecessors. Besides, their sheer scale brings new challenges—the increasing computational cost heightens the barriers to entry to NLP research.

The first part of the talk will discuss innovations in neural architectures that can help address the efficiency concerns of today’s NLP. I will present algorithms that reduce state-of-the-art NLP models’ overhead from quadratic to linear in input lengths without hurting accuracy. Second, I will turn to inductive biases grounded in the inherent structure of natural language sentences, which can help machine learning models generalize. I will discuss the integration of discrete, symbolic structure prediction into modern deep learning.

I will conclude with future directions towards making cutting-edge NLP more efficient, and improving NLP’s generalization to serve today’s language technology applications and those to come in the future.

**Bio:** Hao Peng is a final year PhD student in Computer Science & Engineering at the University of Washington, advised by Noah A. Smith. His research focuses on building efficient, generalizable, and interpretable machine learning models for natural language processing. His research has been presented at top-tier natural language processing and machine learning venues, and recognized with a Google PhD fellowship and an honorable mention for the best paper at ACL 2018.