

# Adapting Transformer models for Document-level Natural Language Tasks

Host: Dragomir Radev



Friday March 10, 2021

4:00 p.m.

Speaker: Arman Cohan

Zoom Presentation

**Abstract:** Transformer models have been extremely effective at producing word- and sentence-level contextualized representations, achieving state-of-the-art results in many Natural Language Processing (NLP) tasks. However, extending these models to document-level NLP tasks faces challenges, including lack of inter-document relatedness information, decreased performance in low-resource settings, and computational inefficiency when scaling to long documents.

In this talk, I will describe a few of my recent works on developing Transformer-based models that target document-level natural language tasks. I will first introduce SPECTER, a method for producing document representations using a Transformer model that incorporates document-level relatedness signals and achieves state-of-the-art results in multiple document-level tasks in the scientific domain. Second, I will describe TLDR, the task of extreme summarization for scientific papers as well as CATTs, a simple yet effective training strategy for generating summaries in low-resource settings. Next, I will discuss the practical challenges of scaling existing Transformer models to long documents, and our proposed solution, Longformer. Longformer introduces a new sparse self-attention pattern that scales linearly with the input length while capturing both local and global context in the document, achieving state-of-the-art results in both character-level language modeling and document NLP tasks. Finally, I'll discuss CDLM, our newly proposed general pre-trained model for addressing multi-document tasks.

**Bio:** Arman Cohan is a Research Scientist at the Allen Institute for AI (AI2). He received his PhD in computer science at Georgetown University in May 2018, advised by Prof. Nazli Goharian. His research is primarily focused on developing general Natural Language Processing capabilities to address information overload, particularly in specialized domains that pose unique challenges. These include core representation learning and language modeling methods, systems and models designed for extracting and summarizing salient information, and models for improved search and categorization in large collections of data.

He has served as program committee member of major NLP venues including ACL, EMNLP, and NAACL in the past 4 years, served as area chair at ACL 2020, ICLR 2021 and NAACL 2021, and organized workshops including SciNLP and SDP.

His research has been recognized with multiple awards including best paper award at EMNLP 2017, area chair favorite paper award at COLING 2018, and Harold N. Glassman Distinguished Doctoral Dissertation award in 2019.