CLSP & CS
VIRTUAL SEMINAR SERIES

Tom McCoy

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Computer Science

"Opening the black box of deep learning: Representations, inductive biases, and robustness"

Monday, January 31, 2022

12:00 PM - 1:15 PM

AMES 234 or
Zoom: https://wse.zoom.us/j/92341914748

ABSTRACT

Natural language processing has been revolutionized by neural networks, which perform impressively well in applications such as machine translation and question answering. Despite their success, neural networks still have some substantial shortcomings: Their internal workings are poorly understood, and they are notoriously brittle, failing on example types that are rare in their training data. In this talk, I will use the unifying thread of hierarchical syntactic structure to discuss approaches for addressing these shortcomings. First, I will argue for a new evaluation paradigm based on targeted, hypothesis-driven tests that better illuminate what models have learned; using this paradigm, I will show that even state-of-the-art models sometimes fail to recognize the hierarchical structure of language (e.g., to conclude that “The book on the table is blue” implies “The table is blue.”) Second, I will show how these behavioral failings can be explained through analysis of models’ inductive biases and internal representations, focusing on the puzzle of how neural networks represent discrete symbolic structure in continuous vector space. I will close by showing how insights from these analyses can be used to make models more robust through approaches based on meta-learning, structured architectures, and data augmentation.

BIOGRAPHY

Tom McCoy is a PhD candidate in the Department of Cognitive Science at Johns Hopkins University. Click here for more information.

HOW TO REACH US

✉ Contactus@cs.jhu.edu
☎ 410-516-8775
🌐 cs.jhu.edu

Johns Hopkins University
Department of Computer Science
3400 N. Charles St | Malone 160
Baltimore, MD 21218