COMPUTER SCIENCE
SEMINAR

Understanding and Generating Human Language

Wei Xu
Georgia Institute of Technology

Abstract: Human language is notoriously complex due to the multitude of ways people can express the same meaning. In this talk, I will present two series of work on machine learning methods to understand the varied expressions in human language and to generate paraphrases for applications, such as reading and writing assistive technology. In the first part, I will showcase how to design learning and ranking models for natural language generation, including a new metric that has been widely adopted as a learning objective and evaluation method. In the second part, I will present new datasets and a class of pairwise neural models for learning textural expressions that convey the same meaning. In contrast to previous work, we focus on extracting paraphrases on a much larger scale and with a much broader range by developing more robust models, leveraging social media data, and crowdsourcing. I will also briefly discuss the connections of my work to computational social science, language and code, and human language instructions. ¡br¿¡br¿

Bio: Wei Xu is an assistant professor in the School of Interactive Computing at the Georgia Institute of Technology. Before joining Georgia Tech, she was an assistant professor at The Ohio State University since 2016. Xus research interests are in natural language processing, machine learning, and social media. Her recent work focuses on language generation, paraphrase acquisition, semantic similarity models for language understanding. She has also worked on crowdsourcing and information extraction. She received her Ph.D. in Computer Science from New York University, B.S. and M.S. from Tsinghua University. She was a postdoctoral researcher at the University of Pennsylvania. She received an NSF CRII Award, a Best Paper Award at COLING 2018, Crowd-Flower AI for Everyone Award, Criteo Faculty Research Award, as well as research funds from DARPA and IARPA. She recently served as a senior area chair for ACL 2020 and an area chair, workshop chair, and publicity chair for EMNLP and NAACL. She co-organizes the Workshop on Noisy User-generated Text annually.

Friday, September 11, 2020, 1:00 pm
https://emory.zoom.us/j/92722816908

COMPUTER SCIENCE
EMORY UNIVERSITY