Representation Learning with Joint Models for Information Extraction

Nanyun (Violet) Peng
John Hopkins University

3/21/2017 at 10:30 am
CoRE A 301

Abstract

There is abundant knowledge out there carried in the form of natural language texts, such as social media posts, scientific research literature, medical records, etc., which grows at an astonishing rate. Yet this knowledge is mostly inaccessible to computers and overwhelming for human experts to absorb. Information extraction (IE) processes raw texts to produce machine understandable structured information, thus dramatically increasing the accessibility of knowledge through search engines, interactive AI agents, and medical research tools. However, traditional IE systems assume abundant human annotations for training high quality machine learning models, which is impractical when trying to deploy IE systems to a broad range of domains, settings and languages. In this talk, I will present how to use deep representation learning to leverage the distributional statistics of characters and words, the annotations for other tasks and other domains, and the linguistic and problem structures, to combat the problem of inadequate supervision, and conduct information extraction with scarce human annotations.

Bio

Nanyun Peng is a PhD candidate in the Department of Computer Science at Johns Hopkins University, affiliated with the Center for Language and Speech Processing and advised by Dr. Mark Dredze. She is broadly interested in Natural Language Processing, Machine Learning, and Information Extraction. Her research focuses on using deep learning for information extraction with scarce human annotations. Nanyun is the recipient of the Johns Hopkins University 2016 Fred Jelinek Fellowship. She has completed two research internships at IBM T.J. Watson Research Center, and Microsoft Research Redmond. She holds a master's degree in Computer Science and BAs in Computational Linguistics and Economics, all from Peking University.

Faculty Host: Ahmed Elgammal and Dimitri Metaxas