

SEMINAR

Compositional Models for Information Extraction

Mark Dredze
Johns Hopkins University

Abstract: Information extraction systems are the backbone of many end-user applications, including question answering, web search and clinical text analysis. These applications depend on underlying technologies that can identify entities and relations as expressed in natural language text. For example, Amazon Echo may answer a user question based on a relation extracted from a news article. A clinical decision support system may offer a physician suggestions based on a symptom identified in the clinical notes from a previous patient visit. In political science, we may seek to aggregate opinions expressed in public comments about a new public policy.

Advances in machine learning have led to new neural models for learning effective representations directly from data that improve information extraction tasks. Yet for many tasks, years of research have created hand-engineered features that yield state of the art performance. I will present feature-rich compositional models that combine both hand-engineered features with learned text representations to achieve new state-of-the-art results for relation extraction. These models are widely applicable to problems within natural language processing and beyond. Additionally, I will survey how these models fit into my broader research program by highlighting work by my group on developing new machine learning methods for extracting public health information from clinical and social media text.

Monday, March 27, 2017, 4:00 pm
White Hall 207

MATHEMATICS AND COMPUTER SCIENCE
EMORY UNIVERSITY