

COMPUTER SCIENCE
SEMINAR

Multi-Structured Inference in Text-to-Text Generation

Kapil Thadani
Columbia University

Abstract: Automated personal assistants and summarization tools are increasingly prevalent in the modern age of mobile computing but their limitations highlight the longstanding challenges of natural language generation. Focused text-to-text generation problems present an opportunity to work toward general-purpose statistical models for text generation without strong assumptions on a domain or semantic representation. In this talk, I will present recent work on a supervised sentence compression task in which a compact integer linear programming formulation is used to simultaneously recover heterogeneous structures which specify an output sentence. This inference strategy avoids cyclic and disconnected structures through commodity flow networks, generalizing over several recent techniques and yielding significant performance gains on standard evaluation corpora.

I will then discuss a number of extensions to this multi-structured generation approach. One line of research explores approximation strategies using Lagrangian relaxation, dynamic programming and linear programming in order to speed up inference while preserving performance. Other extensions exploit the flexibility of the formulation and extend it with minimal additions to new problems such as the more challenging task of merging sentences, as well as to new structures including directed acyclic graphs that represent frame semantics. Finally, I will briefly discuss our use of multi-structured inference in other natural language applications such as summarization and alignment.

Friday, March 21, 2014, 3:00 pm
Mathematics and Science Center: W201

MATHEMATICS AND COMPUTER SCIENCE
EMORY UNIVERSITY