NLP Research Group: MIT

Wuwei Lan, Wei Sun
NLP@MIT

• Introduction
  • group @ CSAIL
  • 2 Professors + 9 Ph.D. + 4 Masters + other undergraduates

• Faculty
  • Regina Barzilay and Tommi S. Jaakkola

• Research Focus
  • very broad: Information retrieval, deep reinforcement learning, recommender systems, Computational biology, Semantic representation and so on.

• Productivity
  • 6~7 top conference papers / year
Regina Barzilay

- Reliable Information Extraction
- Reinforcement learning by acquiring external evidence (EMNLP 2016)
- Interpretable Neural Models
- Rationalizing Neural Predictions (EMNLP 2016)
Tommi S. Jaakkola

Biography
1992, M.S in theoretical physics from Helsinki University of Technology
1997, PhD in computational neuroscience from MIT
1998-now Professor at MIT
Research Synopsis

- **On the theoretical side**
  - statistical inference and estimation
- **On the applied side**
  - NLP, computational biology, recommender, information retrieval
On-going projects and papers

1. **Perturbation models**
   Structured prediction: From gaussian perturbations to linear-time principled algorithms. In Uncertainty in Artificial Intelligence (UIA), 2016

1. **Syntactic and semantic parsing**
   word embeddings as metric recovery in semantic spaces. TACL 2016

1. **Recommender systems**
   Controlling privacy in recommender systems. NIPS 2014

1. **Computational biology**
   Learning population-level diffusions with generative {RNN}s. ICML 2016

1. **Information retrieval/extraction**
What’s interesting?

Topic Modeling in Twitter: aggregating tweets by conversations  ICWSM 2016

1. Background:
Topic Modeling Techniques: Latent Dirichlet Allocation (LDA) and Author-Topic Model (ATM) -> For sufficient long documents with regular vocabulary and grammatical structure

2. what’s about the tweets? (short document and noisy data)
   -> preprocessing tweets for ungrammatical structure and informal language
   -> pooling techniques to aggregate tweets into long documents: User-pooling, Hashtag-pooling and conversation-pooling

3. Can we build a model solve the topic modeling problem in twitter directly?