Invite Speakers for DSTC6 Workshop

Dr. Richard Socher (Salesforce)

Title: Dialogue as multitask learning and question answering

Abstract: Natural language dialogue is the most natural way to interact with a computer. However, in order to work towards natural open ended dialogue we need to tackle a variety of problems that arise in multitask learning. To this end I will outline several challenges and solutions as we try to jointly train a single question answering model to answer a variety of questions.

Bio:
www.socher.org
Richard Socher is Chief Scientist at Salesforce and an adjunct professor at the Stanford Computer Science Department. He leads the company's research efforts and brings state of the art artificial intelligence solutions into the platform. Prior, Richard was the CEO and founder of MetaMind, a startup acquired by Salesforce in April 2016. MetaMind's deep learning AI platform analyzes, labels and makes predictions on image and text data so businesses can make smarter, faster and more accurate decisions. Richard was awarded the Distinguished Application Paper Award at the International Conference on Machine Learning (ICML) 2011, the 2011 Yahoo! Key Scientific Challenges Award, a Microsoft Research PhD Fellowship in 2012, a 2013 "Magic Grant" from the Brown Institute for Media Innovation, the best Stanford CS PhD thesis award 2014, the 2014 GigaOM Structure Award and is currently a member of the World Economic Forum's 'Young Global Leaders' Class of 2017.

Dr. Dhruv Batra (Georgia Tech / FAIR)

Title: Visual Dialog: Towards AI agents that can see, talk, and act

Abstract: We are witnessing unprecedented advances in computer vision and artificial intelligence (AI). What lies next for AI? We believe that the next generation of intelligent systems (say the next generation of Google's Assistant, Facebook's M, Apple's Siri, Amazon's Alexa) will need to posses the ability to `perceive' their environment (through vision, audition, or other sensors), `communicate' (i.e., hold a natural language dialog with humans and other agents), and `act' (e.g., aid humans by executing API calls or commands in a virtual or embodied environment), for tasks such as:

- Aiding visually impaired users in understanding their surroundings or social media content (AI: 'John just uploaded a picture from his vacation in Hawaii', Human: 'Great, is he at the beach?', AI: 'No, on a mountain')
- Aiding analysts in making decisions based on large quantities of surveillance data (Human: 'Did anyone enter this room last week?', AI: 'Yes, 27 instances logged on camera', Human: 'Were any of them carrying a black bag?'),
- Interacting with an AI assistant (Human: 'Alexa – can you see the baby in the baby monitor?', AI: 'Yes, I can', Human: 'Is he sleeping or playing?').
- Robotics applications (e.g. search and rescue missions) where the operator may be 'situationally blind' and operating via language (Human: 'Is there smoke in any room around you?', AI: 'Yes, in one room', Human: 'Go there and look for people').

In this talk, I will present a range of projects from my lab (some in collaboration with Prof. Devi Parikh's lab) towards building such visually grounded conversational agents.

Bio:
https://cc.gatech.edu/~dbatra/files/bio.txt