LINGUISTICS COLLOQUIUM

Will Lewis

Microsoft Research

Microsoft Translator Team

Speech Transcription and Translation in the Classroom and Lecture Setting: Producing "Caption-like" Output

Significant improvements in Automated Speech Recognition (ASR) and Machine Translation (MT) in recent years have brought Speech Translation to the fold as a viable technology. In particular, the application of deep learning has had dramatic impacts on ASR quality; Seide et al 2011, for instance, showed a 32% reduction in Word Error Rate using neural models over the previous state-of-the-art Gaussian models, with no changes in training data, and these improvements have only widened in subsequent years (see Xiong et al 2016, for instance, on achieving human parity error rates on Switchboard). Likewise, the application of deep learning to Machine Translation has shown equally dramatic improvements (e.g., Cho et al 2014, Devlin et al 2014). However, pairing really good speech transcription with really good machine translation does not necessarily generate really good speech translation. People just don't talk in full sentences--input that's generally required for current, state-of-the-art MT systems--people are highly disfluent, and they often use vocabulary that is not recognized (e.g., technical jargon, product names, place and people's names, etc.). Passing a disfluent stream of text produced by ASR to MT generates output that is difficult to interpret at best, and complete gibberish at worst. Our team has done significant work to produce fluent transcripts from disfluent input, using a process we call TrueText. Passing transcripts through TrueText, and then handing these modified transcripts to MT systems that have been tuned to conversational input produces much better results. An added side effect is that the monolingual transcripts are also much easier to read, resembling caption-like output. Further, by modifying ASR on-the-fly by adapting it to the vocabulary that is likely to be used in these scenarios (most notably in classrooms and lectures), we are able to produce transcripts that are much easier to read in the monolingual setting, and likewise much easier to interpret in the bilingual setting. I will explore these technologies, how we have used them in the Microsoft Translator infrastructure, and review the pilots we're currently conducting with university and school partners. I will also do a live demo during my talk that can be broadcast to audience members' own devices in languages of their choice.

Friday, December 1, 2017
3:30 – 5:00 p.m.
Miller 301
Reception will follow in the same location