

CIDSE Invited Talk with Jonathan May

Resource-Constrained Neural Machine Translation



Thursday, January 30, 2020

1:00 PM

BYENG 510

Abstract:

Automated translation between human languages, one of the oldest applications envisioned for computers, has seen a meteoric rise in quality and use over the past decade. However, these gains have not lifted all boats; indeed, most of the world's languages still cannot be translated well by computer. This is because the neural network models responsible for these gains require large amounts of data and computing power to train, limiting their applicability to translation of major world languages and their utility to wealthy institutions. Building new training data for these languages is also hard to find and, due to a lack of access to bilingual speakers, hard to create. In this talk I discuss recent research from our lab that addresses the spectrum of low resource problems in machine translation. Specifically, I will describe our language-universal tools that enable vocabulary sharing, our systems that enable non-speaker humans to generate translation data, and our approaches to transfer learning that enable state-of-the-art translation quality with limited data and computing resources.

BIO:

Jonathan May is a Research Assistant Professor in the Computer Science Department at USC and the Information Sciences Institute (ISI). His research interests include automata theory, machine translation, semantic parsing, dialogue systems, information extraction, and creative generation. He received a PhD from USC in 2010 and BS/MS from UPenn in 2001. He was a co-organizer of the International Workshop on Semantic Evaluation (SemEval) and is the current treasurer of the North American Chapter of the Association for Computational Linguistics (NAACL). He has received a research award from ISI, an outstanding paper award from NAACL, and a best demo paper award from ACL.

Hosted by: Yezhou Yang