Data Analytics with Heterogeneous Supervision

Gerard de Melo
Tsinghua University, Beijing

3/22/2016 at 10:30 am
Core A (Room 301)

Abstract

While unprecedented amounts of data are now readily available, it is often not clear how to exploit these forms of data for machine learning, natural language processing, and artificial intelligence. In this talk, I will describe a series of results on methods that integrate and learn from multiple, possibly quite heterogeneous kinds of data. In natural language processing, for instance, these allow us to produce massively multilingual word representations, covering over 200 languages. Combined with data integration methods, we can also produce representations of entities and large-scale knowledge graphs such as UWN/MENTA, one of the largest multilingual taxonomies, and Lexvo.org, a major hub in the Web of Data. Finally, I will outline new methods for mining from multiple modalities, including images and video, that have the potential to take us towards cognitive computing.

Bio

Gerard de Melo is a Tenure-Track Assistant Professor at Tsinghua University, Beijing, where he is heading the Web Mining and Language Technology group. He has published over 60 research papers in these areas, with Best Paper awards at CIKM 2010, ICGL 2008, and the NAACL 2015 Workshop on Vector Space Modeling, as well as an ACL 2014 Best Paper Honorable Mention, a Best Student Paper Award nomination at ESWC 2015, and the WWW 2011 Best Demonstration Award, among others. Prior to joining Tsinghua, de Melo had spent two years as a Visiting Scholar at UC Berkeley, working in the ICSI AI group. He received his doctoral degree at the Max Planck Institute for Informatics in Germany. For more information, please refer to his home page at http://gerard.demelo.org.