

## Spring 2015 Seminar Series

Presented by the CS Division

### MACHINE LEARNING WITH WORLD KNOWLEDGE

TUESDAY MARCH 3, 2015

10:00 AM – HEC 356

Machine learning algorithms have become pervasive in multiple domains and have started to have impact in applications. Nonetheless, a key obstacle in making learning protocol realistic in applications is the need to supervise them, a costly process that often requires hiring domain experts. However, while annotated data is difficult to get, we have available large amounts of data from the Web. In this talk, I will introduce learning paradigms which use existing world knowledge to “supervise” machine learning algorithms. By “world knowledge” we refer to general-purpose knowledge collected from the Web, and that can be used to extract both common sense knowledge and diverse domain specific knowledge and thus help supervise machine learning algorithms. I will discuss two projects, demonstrating that we can perform better machine learning and text data analytics by adapting general-purpose knowledge to domain specific tasks. For the first project, I will introduce the dataless classification algorithm which requires no labeled data to perform completely unsupervised text classification. In this case, the Wikipedia knowledge is embed to represent the text documents and the category labels into the same semantic space. For the second project, I will discuss how to perform hierarchical clustering of domain-specific short texts, e.g., Web queries and tweets, using a probabilistic concept based knowledge base, Probase. In both cases, we provide realistic and scalable algorithms to address large scale and fundamental text analytics problems.

### DR. YANGQIU SONG

University of Illinois at Urbana-Champaign



Dr. Yangqiu Song is a post-doctoral researcher at the Cognitive Computation Group at the University of Illinois at Urbana-Champaign. Before that, he was a post-doctoral fellow at Hong Kong University of Science and Technology and visiting researcher at Huawei Noah's Ark Lab, Hong Kong (2012-2013), an associate researcher at Microsoft Research Asia (2010-2012) and a staff researcher at IBM Research China (2009-2010) respectively. He received his B.E. and Ph.D. degrees from Tsinghua University, China, in July 2003 and January 2009, respectively. His current research focuses on using machine learning and data mining to extract and infer insightful knowledge from big data. The knowledge helps users better enjoy their daily living and social activities, or helps data scientists do better data analytics. He is particularly interested in working on large scale learning algorithms, on natural language understanding, text mining and visual analytics, and on knowledge engineering for domain applications.

*Hosted by: Dr. Mark Heinrich*

