Department of Data Science

Data Science Seminar Series

Automated Scientific Knowledge Extraction from Massive Text Data

Xuan Wang, Ph.D.
Ph.D. Candidate
The University of Illinois at Urbana-Champaign

Date: Wednesday, February 9th, 2022
Time: 2:30 PM – 3:30 PM EST
Location: Zoom Virtual Room
Web Link: Zoom Meeting Room Link

Text mining is promising for advancing human knowledge in many fields, given the rapidly growing volume of text data (e.g., scientific articles, medical notes, and news reports) we are seeing nowadays. In this talk, I will present my work on automatically extracting scientific knowledge from massive text data to enable and accelerate scientific discovery. First, I will talk about my work on information extraction with minimum human supervision. With the growing volume of text data and the breadth of information, it is inefficient or nearly impossible for humans to manually find, integrate, and digest useful information. To address the above challenge, I have developed methods that automatically extract entity and relation information from massive text data with minimum human supervision. Second, I will talk about my work on literature-based scientific knowledge discovery. This research direction aims to enable and accelerate real-world knowledge discovery with the rich information we automatically extracted from scientific text. I have collaborated with domain experts in various scientific disciplines (e.g., chemistry, biomedicine, and health) to achieve this goal. Last, I will conclude my talk with future directions on using text mining to address open scientific problems, such as to assist chemical and biological molecule design and to support clinical decision-making.

Xuan Wang is a fifth-year Ph.D. student in the Computer Science Department at the University of Illinois at Urbana-Champaign (UIUC). She is working in the Data Mining Group under the supervision of Prof. Jiawei Han. Xuan received M.S. in Statistics (2017) and M.S. in Biochemistry (2015) from UIUC. She received B.S. in Biological Science (2013) from Tsinghua University, China. Her research interests are in Natural Language Processing (NLP) and Data Mining, emphasizing applications to Biological and Health Sciences. Her research has focused on two directions: (1) information extraction with minimum human supervision and (2) literature-based scientific knowledge discovery. Xuan has published about 20 research/demo papers in top NLP conferences (e.g., ACL, EMNLP, and NAACL) and biomedical informatics journals (e.g., Bioinformatics) and conferences (e.g., ACM-BCB and IEEE-BIBM). She is the recipient of the Best Demo Paper Award at NAACL'21 and the YEE Fellowship Award in 2020-2021.