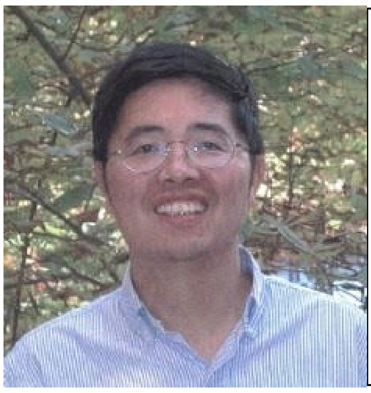
NJIT | INSTITUTE FOR DATA SCIENCE

Department of Data Science

Data Science Seminar Series

Big Data Challenges and Opportunities: Three Case Studies



Xiaohua Hu, Ph.D.
Professor
Drexel University

Date:	Wednesday, January 26th, 2022
Time:	2:30 PM – 3:30 PM EST
Location:	Zoom Virtual Room
Web Link:	Zoom Meeting Room Link

Big Data is transforming science, engineering, medicine, healthcare, finance, business, and ultimately our society itself. In this talk, I will discuss the big data challenges, opportunities and its applications in three case studies with extensive experimental evaluations:

(1) Data analysis and visualization for Microbiome Data: Microbiome datasets are often comprised of different representations or views which provide complementary information to understand microbial communities, such as metabolic pathways, taxonomic assignments, and gene families. Data integration methods based on nonnegative matrix factorization (NMF) combine multi-view data to create a comprehensive view of a given microbiome study by integrating multi-view information. We are presenting a novel variant of NMF called Laplacian regularized joint non-negative matrix factorization (LJ-NMF) for integrating functional and phylogenetic profiles from HMP, and a multiple maps t-SNE regularization method for visualization of mom-metric relationships in microbiome data
(2) Video popularity prediction by sentiment propagation via implicit network: Video popularity prediction is very important in many real applications such as recommendation systems and investment consulting. However, four constraints have limited most existing works' usability. First, most feature oriented models are inadequate in the social media environment, because many videos are published with no specific content features, such as a strong cast or a famous script. Second, many studies assume that there is a linear correlation existing between view counts from early and later days, but this is not the case in every scenario. Third, numerous works just take view counts into consideration, but discount associated sentiments. Nevertheless, it is the public opinions that directly drive a video's final success/failure. Also, many related approaches rely on a network topology, but such topologies are unavailable in many applications. We propose a Dual Sentimental Hawkes Process (DSHP) to cope with all these challenging problems. DSHP's innovations are reflected in three ways: (1) it breaks the ”Linear Correlation” assumption, and implements Hawkes Process; (2) it reveals deeper factors that affect a video's popularity, and (3) it is topology free.
(3) Knowledge Extraction from Massive Scholarly Materials Text Data: The newly NSF funded Institute for Data-Driven Dynamical Research (ID3) aims to address the challenge of predicting dynamical processes in materials, including ion and molecular transport, catalytic pathways, and phase transformations in metamaterials, with a focus on discovering fundamentally new mechanisms and pathways. In this research project, we aim to automatically extract materials knowledge from scholarly text data to accelerate materials discovery. We will discuss various deep learning models in name entity recognition, relation extraction and knowledge graph construction and automatically extract materials knowledge with less manually annotated data.

Xiaohua Tony Hu is a full professor at Drexel University in the College of Computing and Informatics. He is also serving as the IEEE Computer Society Bioinformatics and Biomedicine Steering Committee Chair and IEEE Computer Society Big Data Steering Committee Chair. Tony is a scientist, teacher and entrepreneur. He joined Drexel University in 2002. He founded the International Journal of Data Mining and Bioinformatics (SCI indexed) in 2006. Earlier, he worked as a research scientist in the world-leading R&D centers such as Nortel Research Center, and Verizon Lab (the former GTE labs). In 2001, he founded the DMW Software in Silicon Valley, California.