THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Department of Computing
電子計算學系

Research Centre on Data Science & Artificial Intelligence
人工智能與數據科學研究中心
RC-DSAI

IEEE Computational Intelligence Society

Distinguished Seminar Series on Data Science & Artificial Intelligence

Embedding and Language Modeling for Effective Text Mining

Prof. Jiawei Han
Michael Aiken Chair Professor
Department of Computer Science
University of Illinois at Urbana-Champaign
USA

8 July 2021 (Thu)
10:00 – 11:00 (HKT, UTC+8)
Online via Zoom
Please register at https://polyu.hk/CMwJo or scan the QR code

All are welcome!

Abstract

The real-world big data are largely dynamic, interconnected and unstructured text. It is highly desirable to transform such massive unstructured data into structured knowledge. Many researchers rely on labor-intensive labeling and curation to extract knowledge from such data. Such approaches, however, are not scalable. We vision that massive text data itself may disclose a large body of hidden structures and knowledge. Equipped with pretrained language models and text embedding methods, it is promising to transform unstructured data into structured knowledge. In this talk, we introduce a set of methods developed recently in our group for such an exploration, including joint spherical text embedding, discriminative topic mining, taxonomy construction, text classification, and joint sentiment analysis. We show that data-driven approach could be promising at transforming massive text data into structured knowledge.

About the Speaker

Jiawei Han is Michael Aiken Chair Professor in the Department of Computer Science, University of Illinois at Urbana-Champaign. He received ACM SIGKDD Innovation Award (2004), IEEE Computer Society Technical Achievement Award (2005), IEEE Computer Society W. Wallace McDowell Award (2009), and Japan's Funai Achievement Award (2018). He is a Fellow of ACM and Fellow of IEEE and served as the Director of Information Network Academic Research Center (INARC) (2009-2016) supported by the Network Science-Collaborative Technology Alliance (NS-CTA) program of U.S. Army Research Lab and co-Director of KnowEnG, a Center of Excellence in Big Data Computing (2014-2019), funded by NIH Big Data to Knowledge (BD2K) Initiative.