

Biomedical Literature Mining

Xiaohua Hu

Despite an influx of molecular data in the form of sequences, structure, transcription profiles etc., most of the protein interaction information relevant to cell biology research still exists strictly in the scientific literature which is written in a natural language that computers cannot easily manipulate. Automatically mining and extracting information from biomedical text holds the promise of easily consolidating large amounts of biological knowledge in computer-accessible form. In this talk, we present a novel approach Bio-IEDM (Bio medical Information Extraction and Data Mining) to integrate text mining and predictive modeling to analyze biomolecular network from biomedical literature databases. Our method consists of two phases. In phase 1, we discuss a semi-supervised efficient learning approach to automatically extract biological relationships such as protein-protein interaction, protein-gene interaction from the biomedical literature databases to construct the biomolecular network. In phase 2, we present a novel clustering algorithm to analyze the biomolecular network graph to identify biologically meaningful subnetworks (communities). The clustering algorithm considers the characteristics of the scale-free network graphs and is based on the local density of the vertex and its neighborhood functions that can be used to find more meaningful clusters with different density level. The experimental results indicate our approach is very effective in extracting biological knowledge from a huge collection of biomedical literatures. The integration of data mining and information extraction provides a promising direction for analyzing the biomolecular network.

Xiaohua (Tony) Hu is currently an associate professor and the founding director of the [data mining and bioinformatics lab](#) at the College of Information Science and Technology. He is the now also serving as the IEEE Computer Society Bioinformatics and Biomedicine Steering Committee Chair and the IEEE Computational Intelligence Society Granular Computing Technical Committee Chair (2007-2009). He joined Drexel University in 2002, founded the [International Journal of Data Mining and Bioinformatics](#) in 2006. Earlier, he worked as a research scientist in a few world-leading R&D centers such as Nortel Research Center , GTE labs and HP Labs. In 2001, he founded the DMW Software in Silicon Valley , California . His research ideas have been integrated into many commercial products and applications.

Tony's current research interests are in biomedical literature data mining, bioinformatics, text mining, semantic web mining and reasoning, rough set theory and application, information extraction and information retrieval. He has published more than 140 peer-reviewed research papers in various journals, conferences and books , co-edited 9 books/proceedings. He has received a few awards including the [2005 National Science Foundation \(NSF\) Career award](#), the [best paper award](#) at the 2007 International Conference on Artificial Intelligence , the [best paper award](#) at the 2004 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, the [2006 IEEE Granular Computing Outstanding](#)

[Service Award](#) , and the [2001 IEEE Data Mining Outstanding Service Award](#) . He is the founding editor-in-chief of the [International Journal of Data Mining and Bioinformatics](#) , an associate editor/editorial board member of four international journals ([KAIS](#) , [IJDWM](#) , [IJSOI](#) and [JCIB](#)). His research projects are funded by the National Science Foundation ([NSF](#)), [US Dept. of Education](#) , and the [PA Dept. of Health](#) .

Tony has 8 years solid industry R& D experience and has converted many original research ideas into research prototype systems and eventually into commercial products. From 1994-1998, he was a research scientist in data mining in Nortel Network Research Center , GTE Labs (Verizon Labs) etc. From 1998-2002, he had designed and developed data mining commercial software in various start-up companies (KSP, Blue Martini Software), [KSP](#) was acquired by Exchange Applications for \$52 millions in April 200. In 2001, Tony founded the company DMW software ([D](#) ata [M](#) ining and [W](#) arehousing) in Silicon Valley , California . He has successfully deployed a few data mining products/systems to some Fortune 100 companies such as Chase, Citibank, Sprint for credit fraud detection, e-personalization and customer management systems.