

Thursday, March 23, 2017

Marietta Campus, J Building, room 217 (J 217)/11:00am - 1:30pm

Mingon Kang, Ph.D.

“Large-scale Data Analysis in Bioinformatics”

ABSTRACT: Computational analysis of large-scale data has been grasping great attentions in Bioinformatics. There are increasing numbers of emerging problems that challenge to solve due to incredibly increasing complexity where traditional methods are inapplicable. In particular, current research challenges in Bioinformatics include data integration, reverse engineering from high-throughput multi-omics genomic data, and modeling complex biological systems. In this talk, I'd like to introduce three projects: (1) data integration of multi-omics data for integrative genomic analysis, (2) integrative gene regulatory network inference, and (3) eQTL epistasis in Bioinformatics.

The importance of an integrative genomic study is steadily increasing in an emerging era of various high-throughput genomic data. Mechanisms of human diseases consist of complex interactions of multiple biological processes such as genetic, epigenetic, and transcriptional regulation. The collection of the multiple genomic data that represents the multiple processes is called 'multi-omics data'. The multi-omics data profiled from human disease samples provide comprehensive global snapshots of the diseases. Due to the rapid development of high-throughput technologies, the integrative genomic study using the multi-omics data has been more highlighted than ever. I will introduce our recent works for the integrative genomic study: (1) Multi-Block and Multi-Task Learning (MBMTL) and (2) Multi-Block Bipartite graph Inference (MB2I). MBMTL identifies biomarkers that play important roles in explaining mechanisms of the human diseases from the multi-omics data. MBMTL also takes a multitask problem into account so that we can identify different functions of the mechanisms. Moreover, a novel multi-block bipartite graph and its inference methods, MB2I and SMB2I, for the integrative genomic study not only integrate the multiple genomic data but also incorporate their intra/inter-block interactions by using a multi-block bipartite graph.

Biological network inference is of importance to understand underlying biological mechanisms. Gene regulatory networks describe molecular interactions of complex biological processes. Graph models are mainly used for gene regulatory networks, where nodes and edges represent genes and their regulations respectively. In the most research, the molecular interactions of gene regulatory networks are inferred from a single type of genomic data, e.g., gene expression data. However, gene expression is a product of sequential interactions of DNA sequence variations, single nucleotide polymorphism, copy number variation, histone modifications, transcription factor, DNA methylation, and many other factors. I will introduce our recent proposal, an Integrative Gene Regulatory Network inference method (iGRN) that can incorporate multi-omics data and their interactions in the graph model of gene regulatory network.

Epistasis is the interactions among multiple genetic variants. It has emerged to explain the 'missing heritability' that a marginal genetic effect does not account for by genome-wide association studies, and also to understand the hierarchical relationships between genes in the genetic pathways. The Fisher's geometric model is common in detecting the epistatic effects. However, despite the substantial successes of many studies with the model, it often fails to discover the functional dependence between genes in an epistasis study, which is an important role in inferring hierarchical relationships of genes in the biological pathway. We justified the imperfectness of Fisher's model in the simulation study and its application to the biological data. Then, we proposed a novel generic epistasis model that provides a flexible solution for various biological putative epistatic models in practice. The proposed method enables one to efficiently characterize the functional dependence between genes. Moreover, we suggest a statistical strategy for determining a recessive or dominant link among epistatic eQTLs to enable the ability to infer the hierarchical relationships.



Biography

Dr. Mingon Kang is an assistant professor in the Department of Computer Science at Kennesaw State University. His current research interests include: Computational biology and bioinformatics in Genetics, Neuroscience, and Psychiatry and big data analytics in Multilingual Web and Computer Vision. Especially, he is focusing on developing novel computational methodologies for sparse learning, subspace learning, data integration, and personalized learning. He has published more than 20 research papers in prestigious journals and conferences. Dr. Kang obtained his Ph.D and master degrees from University of Texas at Arlington in 2015 and 2010 respectively, and has a B.E. in Computer Engineering from Hanyang University in South Korea. From 2001 to 2008, he had been a team leader of a start-up company that provides solutions of a personalized web browser and integrated mobile systems.