Abstracts

Hongzhe Li, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine

Pathway-Based Regression Modeling of Genomic Data

High-throughput genomic data provide an opportunity for identifying pathways and genes that are related to various clinical phenotypes, including censored survival phenotypes. Besides these genomic data, another valuable source of data is the biological knowledge about genes and pathways that might be related to the phenotypes of many complex diseases. Databases of such knowledge are often called the metadata. In microarray data analysis, such metadata are currently explored in post hoc ways by gene set enrichment analysis but have hardly been utilized in the modeling step. In this talk, we present two pathway-based regression models, including the pathway-based generalized linear and Cox regression models and the nonparametric pathways-based regression (NPR) models to efficiently integrate genomic data and metadata. Such pathway-based models consider multiple pathways simultaneously and can allow complex interactions among genes within the pathways. We present a pathway-based gradient descent boosting procedure for identifying the pathways that are related to the phenotypes. Our simulation studies indicate that the proposed boosting procedure can indeed identify relevant pathways. Application to a gene expression data set on breast cancer distant metastasis identified that Wnt, apoptosis and cell cycle regulated pathways are more likely related to the risk of distant metastasis among lymph-node-negative breast cancer patients. Results from analysis of other two breast cancer gene expression data sets indicate that the pathways of Metalloendopeptidases (MMPs) and MMP inhibitors, as well as cell proliferation, cell growth and maintenance are important to breast cancer relapse and survival. We also observed that by incorporating the pathway information, we achieved better prediction for cancer recurrence. Extensions to incorporate information contained in the pathway structures in the framework of spectral graph theory and to model the pathway-pathway interactions will also be discussed.