Disease-Pathway Identification with Pathway-Pathway Correlation using Bayesian Inference

Naruemon Pratanwanich* and Pietro Lio
Computer Laboratory, University of Cambridge
*np394@cam.ac.uk

Disease-associated pathway identification is an important step to personalised medicine. In a pathway genes are grouped based on their properties such as their functionality and their physical interactions, and will represent a layer of complexity above the single genes. Since a gene can regulate others in different pathways via a series of actions, it implies that pathways themselves are correlated. To the best of our knowledge, our work is the first effort to consider this issue concurrently with disease-related pathway identification. Moreover, the molecular changes regarding gene expression levels can be measured under various diseased conditions thanks to the advent of microarray technology. Therefore, it is of great interest to incorporate the transcriptomic profiling with pathway-based analysis to explore not only the responsive pathways but also their correlation during disease development. Here, our goal is to simultaneously elucidate the latent disease-related pathways and reconstruct a graph demonstrating the pairwise relationship among these pathways from microarray data. We have developed the mathematical modeling method called matrix factorization under the Bayesian framework by treating pathways annotated by Kyoto Encyclopedia of Genes and Genomes (KEGG) database as a latent space shared by both genes and diseases. We then applied the spike and slab prior, consisting of a point mass at zero (spike) and a normal distribution (slab), to induce the sparsity of the gene-pathway matrix and also introduced the Gaussian distribution with respect to an undirected graph called Gaussian Markov Random Field (GMRF) as a prior imposing on the latent pathway vectors to model the pairwise correlation between pathways. The main novelty yielding from the model is the graph of the pathway dependency. Such a graph structure can serve as a bridge of the hidden mechanisms between disease-disease and even drug-disease in the context of pathway-based activities.

Keywords: Bayesian inference, matrix factorization, gene expression, disease-pathway, pathway-pathway correlation