

Statistical Techniques for Temporal Microarray Data Analysis

Ritesh Krishna

Department of Computer Science, University of Warwick, UK

Recent advances in experimental and computational techniques have revolutionized the field of molecular biology. On the one hand experimental techniques allow us to perform sophisticated and coarse-grained experiments to produce massive amount of observation data, while on the other hand the computational techniques are playing a big role in understanding this data and building hypothesis for understanding of the underlying biological system.

Reconstructing gene-regulatory networks is one of the key problems of functional genomics. A gene network can be visualized as a graph in which each node represents a gene and the interactions between them can be represented by the edges in the graph. The edges can represent direct or indirect connections between the genes. Large-scale monitoring of gene-expression is considered to be one of the most promising techniques for reconstructing gene networks. Technologies like Microarrays are generating abundant amount of data which could be used for reconstructing gene networks. A variety of approaches have been proposed to infer gene-regulatory networks from time-course microarray data, such as Boolean networks, Bayesian networks, difference equations and differential equations etc. These approaches are based on prior biological understanding of the molecular mechanism. Alternatively, models can be constructed solely on analyzing the data itself. Such data driven models allow multivariate biological measurements to reveal new and unanticipated biological insights. There are three important stages in analysis of Microarray data: 1) normalization, to remove the unwanted experimental variations from the data 2) clustering, to reduce the data dimension and 3) network inference, to understand the interactions between observed variables. The present talk is about the application of multivariate data-driven statistical techniques to achieve the three stages in Microarray time series gene expression data analysis. I will talk about the techniques developed to deal with such datasets for Arabidopsis Thaliana plant and the results obtained.