Poster J-54

Identifying Individual Genome-Scale Models: Integrating Reverse Engineering with Classification Analysis

Authors:
Bahrad A. Sokhansanj (*Drexel University*)
Suman Datta (*Drexel University*)
Diane C. Lim (*Drexel University*)

Short Abstract: "Personalized medicine" requires understanding molecular and cellular processes underlying physiological changes and inter-individual differences in human health. We implement a novel method combining large scale discriminant analysis methods with a fuzzy logic scheme for individual-based models for the function of biomolecular networks as they result in health changes.

Long Abstract:

We are developing a platform to develop models for biomolecular network dynamics associated with physical changes and inter-individual differences in human health. Practical applications include finding effective molecular targets for diagnostics, therapy, and individualized therapy (which combines personalized prognostics and therapeutic intervention). High-throughput measurement technologies generate experimental data on the scale of the whole cellular genome and proteome. These large data sets inform the generation of large-scale models of complex systems that regulate cellular response to stimuli and disease progression. Diagnostics research focuses on the value of particular variables, such as the concentration of a molecule in the serum, as a means of identifying a human health state. We are using the discriminant analysis methods from molecular diagnostics in combination with "reverse engineering" methods designed to identify models that are consistent with integrated biological data sets to determine the function of biomolecular networks as they result in health changes.

To deal with the heterogeneity of biological data, and in particular to incorporate qualitative text information and semi-quantitative microarray and proteomic data sets, we use fuzzy logic as a universal method of modeling biomolecular network interactions and dynamics. Our previous work has shown that a consistent scheme of translating gene expression ratios from microarray data to fuzzy sets can be used to accurately and robustly model yeast cell cycle gene networks based on noisy data [1]. In that work, we exhaustively generated all plausible linear fuzzy network models based one experimental data. Now, we are using a combination of fuzzy pre-clustering of correlated variables evolutionary approach to generating plausible models to make the biomolecular network identification problem tractable for data sets including thousands of variables (i.e., genes in a microarray experiment). We have tested our approach on data for cell response to bacterial infection [2] and compared results to two methods of obtaining continuous dynamic network models [3]-[4]. In addition, we have used multivariate discriminant analysis to identify components of the biomolecular network models that change during different biological exposures and among different individuals, demonstrating this approach on the data from [2] as a template. We will also present the analysis of our own gene expression data from the response of patients with chronic

obstructive pulmonary disease (COPD) and non-chronically-diseased patients to acute bacterial infection in the intensive care unit (ICU).

[1] Sokhansanj BA, Fitch JP, Quong JN, Quong AA (2004) Linear fuzzy gene network models obtained from microarray data by exhaustive search. BMC Bioinformatics 5: 108.
[2] Boldrick JC, Alizadeh AA, Diehn M, Dudoit S, et al. (2002) Stereotyped and specific gene expression programs in human innate immune responses to bacteria. Proc Natl Acad Sci USA 99: 972-977.
[3] Guthke R, Moller U, Hoffman M, Thies F, Topfer S (2005) Bioinformatics 21: 1626-1634.
[4] Yeung MK, Tegner J, Collins JJ (2002) Proc Natl. Acad Sci USA 99: 6163-6168.