

Poster J-28
Evidential Reasoning and
Validation for Gene Expression
Data



Authors:

Noppadon Khiripet (*National Electronics and Computer Technology Center, Thailand*)

Short Abstract: Inferring the network structure of genes poses a serious challenge in microarray analysis where the main limitation is the amount of available experimental data. Integrating evidence from the NCBI and KEGG in learning Bayesian network framework helps improve the reliability of clustering as well as gene network inferring procedures.

Long Abstract:

Complicated interactions among genes and gene products result in biochemical interactions and regulatory activities. One goal of system biology study is to recover the true topology of these interaction networks through suitable mathematical models. This process often starts with grouping genes with similar expression patterns in order to identify genes that contribute to common functions and then modeling the states and their changes using graphical models. Recently, graphical models have emerged as useful tools due to the ability to describe the stochastic properties of the association and dependence structures in complex high-dimensional data. However, when complicated graphical models such as Bayesian networks, state-space models, graphical Gaussian models, etc., are applied to gene expression data analysis, their capabilities are limited by the amount of available experimental data. The sample size is small compared to the greater number of considered genes.

In this study, an evidential reasoning framework is presented to address the above problem. To improve the reliability of the candidate genes, evidence from other data sources need to be incorporated, particularly in this study is the data from the KEGG and NCBI's ONIM. The Kyoto Encyclopedia of Genes and Genomes (KEGG) is the database of genomic and molecular information. This information can be used for modeling and simulation of biological systems. On the other hand, the Online Mendelian Inheritance in Man (ONIM) from NCBI represents lists of genes associated with their phenotypes. Unlike standard approach which reconstruct the gene network directly from gene expression data, these two data sources provide meaningful and expert-approved evidence in reconstructing a partial Bayesian belief network structure while the analysis of gene expression data provides another evidence to confirm against the belief network.

As a result, this approach shows its usefulness in providing the reliability checking on the quality and the analysis of gene expression data. It is also possible from the result of the experiment to provide new insights and refine the existing biological pathways and networks