Genetic network architecture identification by conditional entropy analysis

David Correa Martins Jr*, Roberto Marcondes Cesar Jr†, Junior Barrera
Department of Computer Science - IME/USP - http://www.vision.ime.usp.br
Gustavo Goldman
FFCRP/USP

Abstract

A metabolic pathway is a sequence of biochemical reactions mediated by enzymes, constructed by proteins, generated from RNA, produced by gene expression in a genetic network. Besides gene expression is also regulated by proteins. These phenomena pathway constitutes a feedback system and the genetic networks are called regulatory, since they define the pathways. Nowadays, expression levels of thousands of genes can currently be measured simultaneously, thus allowing the observation of different aspects of system dynamics. Eukaryotic cells respond to DNA damage by arresting the cell cycle and modulating gene expression to ensure efficient DNA repair. Saccharomyces cerevisiae MEC1, the human ATR kinase homolog, plays central roles in transducing the damage signal. The procedure below described is being applied to identify genes that belong to the MEC1 hierarquical regulation. We used available microarray data where the genome-wide expression patterns of wild type cells were compared to mutants defectivein Mec1 signaling, under normal growth conditions and in response to the methylating agent metheylmethane sulfonate (MMS) and ioinizing radiation.

A DDS (Discrete Dynamical System) is a finite set of equations that describes the sequential evolution of a vector of discrete variables, called state, under the action of some discrete external forces, called inputs. The next state in time t+1 is computed by a function called transition function, which depends on the states in the previous instants of time, that is, t, t-1, t-2, ... The transition function is decomposed in a vector of functions, called component functions, that compute the transition of each state component.

In this work, we model a genetic network by a DDS, with some random parameters. The state vector is composed by gene expressions, where each gene is a component of this vector. The system architecture is the graph of dependence between genes, i.e. the output and inputs of the component functions. The goal of this work is to present a methodology for finding the architecture of a genetic network by observing time samples of the system dynamics.

Information theory studies random variable dependence by measures such as conditional entropy and mutual information [1]. When the independence of two variables increases, the conditional entropy also increases, since there is less information to concentrate the mass of the conditional probability. In this work, we explore this fact by computing conditional entropy between the gene regulated and the candidate regulatory genes. The entropy of several candidate regulatory genes is computed and organized in a U-type curve. For a given regulated gene, the minimum point of the U-curve defines the regulatory genes.

References

[1] P. D'haeseleer, S. Liang, and Roland Somogyi. Gene expression data analysis and modeling. In *Pacific Symposium on Biocomputing*, Hawaii, January 1999.

*Supported by grant 02/04611-0 from FAPESP - Brazil
†Supported by FAPESP and CNPq - Brazil