Modeling Genetic Networks: Comparison of Static and Dynamic Models

Cristina Rubio-Escudero[1], Oscar Harari[1], Oscar Cordón[1,2], and Igor Zwir[1,3]

[1] Department of Computer Science and Artificial Intelligence,
[2] European Center for Soft Computing, Mieres, Spain.
[3] Howard Hughes Medical Institute, Washington University School of Medicine, St. Louis, MO.
{crubio, oharari, ocordon, zwir}@decsai.ugr.es

Abstract. Biomedical research has been revolutionized by high-throughput techniques and the enormous amount of biological data they are able to generate. The interest shown over network models and systems biology is rapidly raising. Genetic networks arise as an essential task to mine these data since they explain the function of genes in terms of how they influence other genes. Many modeling approaches have been proposed for building genetic networks up. However, it is not clear what the advantages and disadvantages of each model are. There are several ways to discriminate network building models, being one of the most important whether the data being mined presents a static or dynamic fashion. In this work we compare static and dynamic models over a problem related to the inflammation and the host response to injury. We show how both models provide complementary information and cross-validate the obtained results.

1 Introduction

Advances in molecular biology and computational techniques permit the systematical study of molecular processes that underlie biological systems (Durbin *et al*., 1998). One of the challenges of this post-genomic era is to know when, how and for how long a gene is turned *on/off*. Microarray technology has revolutionized modern biomedical research in this sense by its capacity to monitor the behavior of thousands of genes simultaneously (Brown *et al*., 1999; Tamames *et al.*, 2002). The reconstruction of genetic networks is becoming an essential task to understand data generated by microarray techniques (Gregory, 2005). The enormous amount of information generated by this high-throughput technique is raising the interest in network models to represent and understand biological systems.

Systems biology research arises at this point as the field to explore the life regulation processes in a cohesive way making use of the new technologies. Proteins have a main role in the regulation of genes (Rice and Stolovitzky, 2004), but unfortunately, for the vast majority or biological datasets available, there is no information about the level of protein activity. Therefore, we use the expression level of the genes as an indicator of the activity of proteins they generate.

Gene networks represent these gene interactions. A gene network can be described as a set of nodes which usually represent genes, proteins or other biochemical entities. Node interaction is represented with edges corresponding to biologic relations.

E. Marchiori, J.H. Moore, and J.C. Rajapakse (Eds.): EvoBIO 2007, LNCS 4447, pp. 78 – 89, 2007.

There is a wide range of models available to build genetic networks up. One of the differences between such models is whether they represent static or dynamic relations. Static modeling explains causal interactions by searching for mutual dependencies between the gene expression profiles of different genes (van Someren *et al.,* 2002). Clustering techniques are widely applied for static genetic network, since they group genes that exhibit similar expression levels.

In dynamic modeling, the expression of a node A in the network at time t_{+1} can be given as the result of the expression of the nodes in the network with edges related to A at time t (van Someren *et al.,* 2002). The understanding of the relations helps to describe all the relations occurring in a given organism we would be able to know the behavior of such organism throughout time.

The question arises as which network model is the most appropriate given a set of data. In the present work we have applied both static (*K*-means clustering method, (Duda and Hart, 1973)) and dynamic network models (a Boolean method, described in (D'onia *et al.*, 2003) and implemented in (Velarde, 2006) and a graphic Gaussian method (GGM) (Schäfer and Strimmer, 2005)) to a set of data derived from an experiment on inflammation and the host response to injury (Calvano *et al*., 2005). The results show how dynamic models are capable to recover temporal dependencies that static models are not able to find. Temporal studies are becoming widely used in biomedical research. In fact, over 30% of published expression data sets are time series (Simon *et al*., 2005).

2 Problem Description

In this work we compare the behavior of static vs. dynamic modeling in a problem derived from the inflammation and the host response to injury. On the one hand, static modeling searches for relations between the expression levels of genes throughout time. The relation found by static methods might not only be similar behavior throughout time (direct correlation), but an inverse correlation (two genes having exactly opposite profiles over time), a proximity on the expression values (distance measures such as Euclidean Distance or City block distance) (see Fig. 1). On the other hand, dynamic modeling retrieves temporal dependencies among genes, i.e., it detects dependencies of a gene at time t_{+1} related to some other(s) gene at time t (see Fig. 1).

To compare the performance of these two models, we have applied them to a data set derived from an experiment over inflammation and the host response to injury as part of a Large-scale Collaborative Research Project sponsored by the National Institute of General Medical Sciences (www.gluegrant.org) (Calvano *et al*., 2005). Human volunteers have been treated with intravenous endotoxin and compared to placebo, obtaining longitudinal blood expression profiles. Analysis of the set of gene expression profiles obtained from this experiment is complex, given the number of samples taken and variance due to treatment, time, and subject phenotype. The data were acquired from blood samples collected from eight human volunteers, four treated with intravenous endotoxin (i.e., patients 1 to 4) and four with placebo (i.e., patients 5 to 8). Complementary RNA was generated from circulating leukocytes at 0, 2, 4, 6, 9 and 24 hours after the and hybridized with GeneChips® HG-U133A v2.0 from Affymetrix Inc., which contains 22216 probe sets, analyzing the expression level of 18400 transcripts and variants, including 14500 well-characterized genes.