IRIS: A Probabilistic Approach to Infer Regulatory Relations in Gene Networks

Morganella S[2], Ceccarelli M[1,2], Zoppoli P[2]

[1] Research Centre on Software Technologies - RCOST, University of Sannio, Benevento, Italy
[2] Biogem s.c.a.r.l., Ariano Irpino (Avellino), Italy

Motivation

In the last years the microarray technology has revolutionized the fields of genetics, biotechnology and drug discovery. The development of efficient algorithms to extract gene networks by high-throughput gene expression data is a major issue in current bioinformatics research. The methods that solve this task are called "reverse engineering algorithms". A gene network, or gene regulatory network, is a collection of genes interacting each other to regulate the activities that underlie the major functions of living cells. Gain knowledge about the gene regulatory networks is very important to predict the response of the systems to external perturbations, such as how identify the target genes of a compound, and others. Our approach tries to solve the task of reverse engineering of regulation function using a given topological description of the network and gene expression profile. The developed method seems to be an important alternative and improvement to the methods of refinement of regulation functions proposed in literature. At URL http://bioinformatics.biogem.it a Matlab implementation of IRIS is available.

Methods

Regulatory rules into gene networks enables the researchers to discover the complex interactions among genes. We propose a new approach to solve the reverse engineering of regulatory relations into gene networks from data provided by microarray technology. The proposed algorithm, called Inference of Regulatory Interaction Schema (IRIS), uses an iterative approach to map the gene expression profile values (both steady-states and time-courses) into discrete states so we then use a probabilistic method to infer the regulation functions of the network, these interaction rules are integrated into a factor graph model. In our approach we assume that a gene may attain two states: active and inactive to distinguish the response of a gene to a given experimental condition, this assumption is commonly used in literature. The IRIS discretization method is composed of a two-level approach, where first we fix the states that have an high likelihood to represent active/inactive states and then we use these states into a relaxation iterative step to try to recover other states. The discrete states obtained by the discretization step are used by IRIS to compute the regulation functions of the gene network providing two different descriptions for each of them: the description in which each interaction is described as a conditional probability table and a description in which each regulation is a truth table. These two different kinds of descriptions allow biologists to perform different analysis on the model, so an inference engine can be used to compute the posteriori probabilities of hidden genes given observed genes by the conditional probability tables of the model, whereas we can use the true tables to find the steady states of the gene regulatory network. In these next analysis steps we deal with the problem of the cyclic structure that a gene network may have, so we use an approach based on the factor graph model to run the inference engine, and we use the concept of feedback set to compute the steady states.

R

In order to validate the proposed method we use both synthetic data and data from microarray experiments. For the former case, we test IRIS on two synthetic networks: the first test shows that IRIS discretization approach has a good accuracy and second test demonstrates that IRIS provides a fast and accurate method to infer the regulatory relations where the Expectation Maximization Maximum a Posteriori (EM-MAP) algorithm is used as comparison method. Moreover, we use our approach to learn the regulation functions of the cell-cycle of the yeast Saccharomyces cerevisiae from microarray expression profiles and the results are compared with the literature, showing that IRIS has an high propensity to provide regulatory relations which are more similar to the real features of the network. Specifically, IRIS has an accuracy about of 80% and using its results we found well-known features studied in literature, among them two well-known steady states of the yeast cell-cycle are inferred.

Contact : morganella@biogem.it