Precedence temporal networks for the representation of temporal relationships in gene expression data

Sacchi L, Larizza C, Magni P, Bellazzi R

Dipartimento di Informatica e Sistemistica, University of Pavia, Pavia

Motivation

The possibility of a genome-wide measure of gene expression offered by the DNA microarray technology is stimulating researchers in dealing with complex scientific challenges. Among them, the reconstruction of gene regulatory networks is particularly interesting. While in functional genomics genes are analyzed as single units to determine new functional information, in gene network reconstruction the interactions between genes are considered in order to infer regulatory connections between single genes or groups of them. In the field of discovering regulatory pathways, research is mainly focused on the analysis of gene expression time series collected by repeating DNA microarray experiments at different points in time. Many approaches based on different techniques of gene interaction representation have been proposed in the literature to deal with the topic of reconstructing gene regulatory networks. The observation that the dynamics of biochemical reactions may not be exactly captured by the (low) sampling time available in DNA microarray experiments has recently led to the introduction of the so-called module networks, where patterns of synchronized gene expressions are introduced. The present work introduces Precedence Temporal Networks (PTN), a novel method to extract from data and graphically represent temporal relationships between genes. Precedence Temporal Networks are a special kind of temporal network, where nodes (genes) are represented by properly defined temporal events while edges identify temporal relationships between the nodes.

Methods

The proposed approach develops through three steps: significant events are first identified in the time series by exploiting a qualitative representation of the profiles based on the technique of Temporal Abstractions (TAs). Precedence and synchronization relationships between abstractions are then searched through a set of properly extracted temporal rules; the resulting relationships are finally mapped into a labeled graph, the Precedence Temporal Network. In more detail, a first simple qualitative representation of the profiles made up of a set of consecutive basic trend TAs is obtained by processing raw data with a suitable algorithm. This description constitutes the starting point for the creation of the complex temporal events, i.e. complex abstractions which describe specific interesting temporal behaviors (typically user-defined) occurring in the data. A complex temporal event holds over an interval and is labeled through a set P={p1,…,pn}, where each pi represents an interesting qualitative behavior and is made up by the composition of simple labels of the kind Increasing, Decreasing, Steady (e.g. Increasing-Decreasing). Precedence and synchronization relationships are then selected through a method for temporal rules extraction that works looking for both the members of the rule coming from the set of the complex temporal events. In a PTN each gene complex behavior will be described by a node; edges will represent the relationships between the nodes extracted by the temporal rules. To specify both synchronization and precedence relationships we distinguish between two types of connections: the first, called cooccurrence connection, links elements characterized by the simultaneity of their temporal events, while the second, called precedence connection, corresponds to a precedence temporal relationship. The result will be a graphical representation of temporal relationships of the kind "an increasingdecreasing pattern in Gene A PRECEDES an increasing-steady pattern of Gene B".

Results

The method has been applied to the analysis of the expression of a subset of genes involved in human cell cycle regulation; in particular, we have analyzed 20 time series of 47 samples corresponding to well-studied genes known to show an expression peak in specific phases of the cell cycle. The extracted network is able to clearly highlight the different cell cycle phases, and in

particular the strong synchronization occurring at phase G1/S boundary and the weaker synchronization of the latter phases. The proposed approach is a novel contribution in the visualization of temporal relationships between a set of variables and its application in the analysis of gene expression data seems useful to summarize cellular behavior. Our approach is different from other methods, since it is devoted to describe precedence and synchronization of gene temporal patterns in a data set. This technique is still under development; future work will be mainly focused on the study of the formal properties of the network, and its application to more complex biological problems.

Contact email: lucia.sacchi@unipv.it