Abstract

The reconstruction of Gene Regulatory Networks (GRNs) from microarray gene expression dataset is of high interest and a very challenging computational problem for a better understanding of the highly complex regulatory mechanisms in the cellular systems of the living organisms.

In the proposed work, we introduce a non-supervised technique to infer GRNs from microarray time-series dataset. The co-expressed genes are initially grouped together using ensemble biclustering technique followed by inferring a Relevance modelling of a GRN. Using the biclustering approach prior to inferring GRN overcomes a major problem of microarray time-series gene expression dataset i.e. the problem of fewer number of time points as compared to the relatively very large number of genes. Moreover the ensembled biclustering approach improves the performance of the biclustering algorithms.

We have used a real microarray time-series dataset of Saccharomyces Cerevisia i.e. yeast with 4471 genes and 5 time points. Firstly, two different time-series based biclustering algorithms namely CCC Biclustering and e-CCC Biclustering are run separately on microarray time-series data to get contiguous column coherent biclusters. For each CCC bicluster, we identify the corresponding e-CCC bicluster based on GO annotations. Then the identified strongly connected biclusters pairs are ensembled. These ensembled biclusters are validated using p-value, Q-value and percentage of gene coverage w.r.t. the annotated genes in the genomes participating in particular molecular functions. For each of the CCC, e-CCC and ensembled biclusters, a subnetwork is inferred in Cytoscape platform using the CYNI tool and ARANCE as the inferring algorithm. Then using an edge selection technique the edges of the three similar subnetworks inferred from the CCC biclusters, e-CCC biclusters and the ensembled biclusters are considered to construct a single subnetwork. Eventually, we construct the GRN by integrating all the subnetworks.

Keywords: · Gene Regulatory Network · Microarray Time-Series Data · Co-expressed genes and coherent patterns · Biclustering · CCC bicluster · e-CCC bicluster· Ensembled bicluster · GO Annotations · Relevance network · ARANCE.