Abstract

Background: The computational reconstruction of Gene Regulatory Networks (GRNs) using different techniques have encountered the challenge of constructing large network because of many parameters to be fitted and the nature of the input data. In fact, contemporary works on GRN inference that involve the use of hybridized techniques especially Artificial Neural Network (ANN) with meta-heuristic optimization techniques have to trade off computational cost for accuracy in reconstructing large-scale GRN. This work designed an efficient feature selection algorithm with GRN model to overcome the dimension problem of input data using biological prior knowledge of co-expression and network sparseness, so as to capture and represent the actual interrelationship among genes.

Methodology: The GRN model is an ensemble Multi-Layer Perceptron (MLP) network incorporating a novel feature selection algorithm termed Fuzzified Adjusted Rand Index (FARI). FARI is developed to investigate and establish the expression trends of genes in an expression profile data. A rank matrix of all genes produced by FARI shows their co-expression
relationship, which is used to co-ordinate the selection of potential predictors as input features into the inference model. Each target gene is modeled separately by updating its parameters independently as several sub-problems of the overall network. The performance of the model is subjected to synthetic, ecoli and Mtb data.

Result: The result indicated an improved accuracy in the construction of large-scale GRN including a significant speed-up. The result on Mtb identified CCL5 as the first expressed gene, which is the same with CCL1 identified by the experimental method. Some of the expressed genes were validated through their biological pathways showing immune responses and host susceptibility to TB.

Conclusion: The included prior biological knowledge in MLP model provided the construction of an accurate large-scale GRN by reducing the potential large search space of GRN modeling. Besides, the model produced two major biological networks from the same process using the same dataset for appropriate biological validation.

References


**Index Terms**

Computer Science Networks
Keywords

Gene Regulatory Network, Multi-Layer Perceptron, Fuzzified Adjusted Rand Index, prior knowledge, co-expression, rank matrix