

# A Multiobjective Phenomic Algorithm for Inference of Gene Networks

Rio G.L. D'Souza<sup>1</sup>, K. Chandra Sekaran<sup>2</sup>, and A. Kandasamy<sup>2</sup>

<sup>1</sup> St Joseph Engineering College, Mangalore, India

<sup>2</sup> National Institute of Technology Karnataka, Surathkal, Mangalore, India  
{rio,kchnitk}@ieee.org, kandy@nitk.ac.in

**Abstract.** Reconstruction of gene networks has become an important activity in Systems Biology. The potential for better methods of drug discovery and of disease diagnosis hinge upon our understanding of the interaction networks between the genes. Evolutionary methods are proving to be successful in such problems and a number of such methods have been proposed. However, all these methods are based on processing of genotypic information. We have presented an evolutionary algorithm for reconstructing gene networks from expression data using phenotypic interactions, thereby avoiding the need for an explicit objective function. Specifically, we have also extended the basic phenomic algorithm to perform multiobjective optimization for gene network reconstruction. We have applied this novel algorithm to the yeast sporulation dataset and validated it by comparing the results to the links found between genes of the yeast genome at the SGD database.

**Keywords:** Gene networks, Phenomic algorithm, Multiobjective optimization, Evolutionary algorithms, Yeast Sporulation, Microarray data analysis.

## 1 Introduction

Advances in methods of gene expression measurement have heralded the advent of high throughput methods such as microarray technology. Biologists now can study hundreds of genes at a time, and such studies lead to the elucidation of relationships between genes which ultimately lead to a better understanding of the cellular processes that form the basis of life. However the datasets that result from such studies have high dimensionality. The challenge is to analyze such datasets without compromising their information content. Several researchers have developed methods of analysis which can determine useful patterns from the datasets without compromising the dimensionality [1].

Gene networks represent relationships between genes, based on observations of how the expression level of each gene affects the expression levels of the others [2]. The determination of these relationships from gene expression measurements is a reverse engineering or reconstruction activity [3]. Evolutionary methods have been found to be useful [4] to analyze and capture the relationships between hundreds of genes. The application of new ideas of evolutionary optimization to the inference of

gene networks is an ongoing process and many non-conventional methods have shown remarkable success [5]. The Phenomic Algorithm, introduced in [6], and further studied in [7], is one such method. It presents an evolutionary approach based on phenotypic interactions rather than genotypic mechanisms which are used in traditional evolutionary algorithms.

In this paper, we have modified the basic phenomic algorithm to handle multiple objectives. It is possible to employ multiobjective optimization to elucidate gene networks which are more biologically plausible [8]. We have used non-dominated sorting in order to determine the pareto-optimal solutions that best represent the balance between the objectives that we have chosen to optimize. We have applied the multiobjective phenomic algorithm to the yeast sporulation dataset [9] and results show a marked improvement in the quality of networks discovered.

The rest of this paper is organized as follows: In Section 2, we review the related work done by others. We devote Section 3 to a discussion about the methodology adopted by the basic phenomic algorithm and its implementation. We discuss the rationale for modification of the basic phenomic algorithm in Section 4 and its actual implementation in Section 5. Finally, Section 6 presents the results and validation, followed by Section 7 which concludes the paper.

## 2 Related Work

While early methods for reconstruction of gene networks focused on inferring Boolean networks [10] others have used differential equations [11], [12], [13] and Bayesian networks [14], [15] to infer qualitative, as well as quantitative models of gene networks. Given that gene networks are intrinsically nonlinear and dynamic systems, some researchers [8], [16] have used the S-system proposed by Savageau [17] in order to formulate an objective function for the evolutionary algorithm that they use to reverse engineer gene networks.

State space models [18] and information theoretic approaches [19], [20] have also been successfully applied to the problem of inferring gene networks from microarray data. In recent years machine intelligence based approaches [21], [22], [23] are becoming popular in this area due to their relative ease of application. A number of multiobjective evolutionary algorithms (MOEAs) have been applied to the problem of reconstructing gene networks from expression data [24], [25]. Notable among these algorithms is the non-dominated sorting genetic algorithm (NSGA) and its variations which have been applied to the problem of classification of cancer based on gene expression data [26], [27], [28].

The application of MOEAs to the elucidation of gene networks is an area which is receiving a large amount of focus from researchers due to the perceived benefits in applications such as drug discovery and the diagnosis of chronic diseases. This has been the motivation for the development of the phenomic algorithm [6], [7] which attempts to solve the problem of requiring an explicit fitness function for the optimization process. In this paper, we extend this algorithm to perform multiobjective optimization.