A Phenomic Approach to Genetic Algorithms for Reconstruction of Gene Networks

Rio G.L. D'Souza[1], K. Chandra Sekaran[2], and A. Kandasamy[2]

[1] St Joseph Engineering College, Vamanjoor Post, Mangalore 575028, India
[2] National Institute of Technology Karnataka, Surathkal, Mangalore 575 025, India
`rio@ieee.org, kchnitk@ieee.org, kandy@nitk.ac.in`

Abstract. Genetic algorithms require a fitness function to evaluate individuals in a population. The fitness function essentially captures the dependence of the phenotype on the genotype. In the Phenomic approach we represent the phenotype of each individual in a simulated environment where phenotypic interactions are enforced. In reconstruction type of problems, the model is reconstructed from the data that maps the input to the output. In the phenomic algorithm, we use this data to replace the fitness function. Thus we achieve survival-of-the-fittest without the need for a fitness function. Though limited to reconstruction type problems where such mapping data is available, this novel approach nonetheless overcomes the daunting task of providing the elusive fitness function, which has been a stumbling block so far to the widespread use of genetic algorithms. We present an algorithm called Integrated Pheneto-Genetic Algorithm (IPGA), wherein the genetic algorithm is used to process genotypic information and the phenomic algorithm is used to process phenotypic information, thereby providing a holistic approach which completes the evolutionary cycle. We apply this novel evolutionary algorithm to the problem of elucidation of gene networks from microarray data. The algorithm performs well and provides stable and accurate results when compared to some other existing algorithms.

Keywords: Genetic algorithms, Phenomic algorithms, Reconstruction, Fitness function, Multiobjective evolutionary algorithms, Gene networks.

1   Introduction

Genetic Algorithms have been used to solve a wide range of search and optimization problems ever since their introduction by Goldberg [1]. The robust and multimodal nature of these algorithms makes them ideally suited for many applications. However, genetic algorithms impose a few pre-requisites which need to be satisfied before applying them to solve a problem. Encoding of the solution into genetic information, development of a fitness function, or objective function, which evaluates each solution, selection of appropriate mechanisms for crossover and mutation, are some of these pre-requisites. Among these, the need for a fitness function is the major hurdle which prevents the potential application of genetic algorithms to many problems.

Reconstruction problems are encountered in reverse engineering scenarios where the model of the system is unknown and needs to be recreated solely from the interactions of the system with the outside world. Generally, a series of stimuli are applied to the system and the response is recorded. The stimulus-response pairs are then utilized to reconstruct the system model. For example, a series of input signals can be applied to an electronic device and the response waveform can then be analyzed to understand the internal make-up of the device.

When applying genetic algorithms to reconstruction problems, the ready availability of stimuli-response pairs can be exploited to avoid the need to develop the fitness function. Each individual of the population can embed one of the stimuli-response pairs, thereby avoiding the need to explicitly evaluate the individual using a fitness function. This novel idea has motivated us into developing a Phenomic approach to the survival-of-fittest phase in a typical genetic algorithm. We call it the phenomic approach since the emphasis is on the phenotype, rather than the genotype. The phenotype is the key player in the environment where survival-of-the-fittest determines which individuals are selected for propagating their genes into the next generation.

We have developed the phenomic approach which simulates an environment where individuals meet and interact. The interactions are modeled to represent the survival-of-the-fittest as it occurs in nature. Since each individual has its response embedded within itself, the selection of some individuals over the others will depend on this response. Thus there is no need to explicitly evaluate each individual using a fitness function. The rest of the genetic algorithm proceeds as usual, with crossover and mutation being used to create a new offspring population, which then becomes the basis for the next generation.

We have implemented the phenomic approach in a genetic algorithm that elucidates gene networks from microarray data. This is a typical reverse engineering problem which seeks to determine the causal gene network that can best account for observed measurements of RNA concentrations made under either different experimental conditions, or at different points of time.

The rest of the paper is organized as follows: in section 2, we review related work done by others in the area of genetic algorithms, especially when applied to reconstruction problems and reverse engineering of gene networks. In section 3, we present details of our approach followed by application of the novel approach to the problem of reconstructing gene networks from microarray data. We then present the results of our experiments, in section 4, and a discussion of its relative strengths and weaknesses. Finally, in section 5, we conclude the paper, with suggestions for possible future directions of research in this area.

2   Related Work

Genetic Algorithms have come a long way in the field of search and optimization. In the following sub-sections, we discuss only those works that are directly related to our work. We discuss work done by others on reconstruction problems, specifically with