Multi-Objective Evolutionary Algorithm for Identifying the Important Parameters of a Complex System

Brajesh Varshney
Research Scholar
ABV-IIITM
Gwalior-474010

Pankaj Srivastava
Associate Professor
Department of Physics
ABV-IIITM, Gwalior-474010

ABSTRACT

In recent decades, considerable research efforts have been devoted to using machine learning or data mining techniques to automatically discover the parameters in multi-objective functions. Among these techniques, the genetic algorithms have been recognized to be particularly powerful in multi-objective problems.
Genetic algorithms play a significant role, as search techniques for handling complex spaces in many fields. These algorithms are based on the underlying genetic process and are optimization algorithms based on the mechanics of natural genetics and natural selection.
Initially, the search space solutions are coded using the binary alphabet for a discrete search space. Even though the underlying objective function is a continuous function, genetic algorithms convert the search space into discrete set of points. In order to obtain the optimum point with a desired accuracy, strings of sufficient length need to be chosen.
Genetic algorithms have also been developed to work directly with continuous variables (instead of discrete variables). In such genetic algorithms, binary strings are not used. Instead, the genes of chromosome are represented as real numbers directly. In such algorithms the solutions are very close to the natural formulation.
In this paper, an Evolutionary Algorithm is developed for identifying the important parameters essential for the multi-objective problem. The main three operators- reproduction, cross-over and mutation are used to create new population of points. The new population is further evaluated and tested for termination criterion. If the termination criterion is not met, the population is iteratively operated by the above three operators and evaluated. This procedure is continued until the termination criterion is met. In reproduction operation, the rank-based elitism roulette wheel selection scheme is adopted. In cross-over operation, a specific cross over operator for this problem is used. In mutation operation, the random substitution method is used.
A computer program for this Evolutionary Algorithm is developed for identifying the important parameters essential for the multi-objective problem. This algorithm is tested by number of cases.

Keywords

Evolutionary Algorithm, Multi-objective problems, Roulette wheel selection scheme

1. INTRODUCTION

Real-world optimization problems often involve a number of characteristics, which make them difficult to solve up to a required level of satisfaction. Those characteristics are [1]-

- Existence of mixed type of variables(such as Boolean, discrete, integer and real)
- Existence of non-linear constraints
- Existence of multiple conflicting objectives
- Existence of multiple optimum solutions
- Existence of stochasticities and uncertainties in describing the optimization problem

In recent decades, considerable research efforts have been devoted to using machine learning or data mining techniques to automatically discover useful medical knowledge and rules [2-3]. Among all these techniques, the genetic algorithms have been identified as powerful in medical diagnosis. Genetic algorithms mimic the principles of natural genetics and natural selection to constitute search and optimization procedures.
In this paper, an Evolutionary Algorithm is developed for identifying the important parameters essential for the multi-objective problem.
In section 2, the multi-objective genetic algorithm is discussed. In section 3, developed computer program for this algorithm is discussed. The conclusions are summarized in section 4.

2. MULTI-OBJECTIVE GENETIC ALGORITHM

Initially, in the genetic algorithms, the search space solutions are coded using the binary alphabet for a discrete search space. Even though the underlying objective function is a continuous function, genetic algorithms convert the search space into discrete set of points. In order to obtain the optimum point with a desired accuracy, strings of sufficient length need to be chosen.

Genetic algorithms have also been developed to work directly with continuous variables (instead of discrete variables). In such genetic algorithms, binary strings are not used. Instead, the genes of chromosome are represented as real numbers directly. In such algorithms the solutions are very close to the natural formulation.
The algorithm will be stopped when the number of iterations exceeds a threshold. It will also be terminated when all the instances of data base is recognized.

2.1 Generation of population

The evolution usually starts from a population of randomly generated individuals and happens in generations. The population of abstract representations is called chromosomes [4]. The randomly selected chromosome is represented by an array of matrix as follows-

$$X_i = [x_{pq}]$$

Where, i = 1 to m randomly selected chromosome
p = number of parameters

q = Number of objectives

The value of X_{pq} varies between -1 to +1.

2.2 Evaluate the Chromosome (Fitness Function)

Each chromosome in the population has an associated fitness to determine which chromosomes are used to form new ones in the competition process, which is called *selection or reproduction* [5]. In multi-objective problem, it corresponds to the number of correct identification of the given number of objectives over the whole data set.

The fitness value of a chromosome

$$F(X_i) = \sum_{i=1}^{d} (IF)_i / N$$

where,
N = Total number of dataset
IF = Indication function

Data set is represented by-

$d_j = [\, d_1, d_2, d_3, \text{-----------}, d_p]$, where j = 1 to N

Correct identification is given by-

$$\text{Maximum}\{ \sum_{t=1}^{N} d_{tj} * x_{pqt} \}$$

If this correct identification = real object of data set at that instance

IF=1
Otherwise IF = 0

2.3 Genetic Algorithm operations

After calculating the fitness value of all the chromosomes of the population as per the method mentioned above, the population is then operated by three main operators- ***reproduction, crossover and mutation***- to create a new population of points. The new population is further evaluated and tested for termination. If the termination criterion is not met, the population iteratively operated by the above three operators and evaluated. This procedure is continued until the termination criterion is met. One cycle of these operations and the subsequent evaluation procedure is known as a generation in genetic algorithm terminology.

2.3.1 Reproduction operation

In reproduction or selection stage, good chromosomes from the population are selected and a mating pool is formed. In this above average chromosomes are picked from the current population and their multiple copies are inserted in the mating pool. The commonly used reproduction operator is the proportionate reproduction operator where the chromosome is selected for the mating pool with a probability proportional to its fitness. To implement this scheme, a **roulette-wheel strategy [6]** is used.

The probability of selection of a i^{th} chromosome is given by-

$$p_i = F_i / \sum_{j=1}^{n} F_j$$

n is a population size

2.3.2 Crossover operation

Since the real encoding is adopted in this study, the standard crossover operation for the binary encoding method cannot be used. Therefore, new and efficient crossover operators [6] have been designed so that search along variable is also possible.

Let us consider $C_i^{(j)}$ and $C_i^{(k)}$ values of design variables C_i in two parent chromosomes j and k. The crossover between these two values will produce the following new value-

$$C_i^{new} = (1- \lambda)* C_i^{(j)} + \lambda * C_i^{(k)}$$

The parameter λ is a random value between zero and one and the operator "*" denotes the element by element matrix multiplication.. This equation calculates a new values of chromosomes $C_i^{(j)}$ and $C_i^{(k)}$. This calculation is performed for each chromosome participating in crossover operation, under certain crossover probability.

2.3.3 Mutation Operation

A crossover operator is mainly responsible for the search of new chromosomes, even though a mutation operator is also used for this purpose. The need for mutation is to create a point in the neighborhood of the current point thereby achieving a local search around the current solution. The mutation is also used to maintain diversity in the population.

In this problem, the random substitution method for mutation operation is used. That is chromosome chosen to mutate is replaced randomly by a new chromosome with certain mutation probability.

3. COMPUTER PROGRAM

A computer program is developed to find out the best chromosome and the important parameter for the multi-objective evolutionary algorithm. This program is developed with Wamp Server and Server Configuration Apache Version: 2.2.17 PHP: Version :5.3.5 and showed in fig. 1 to 3.

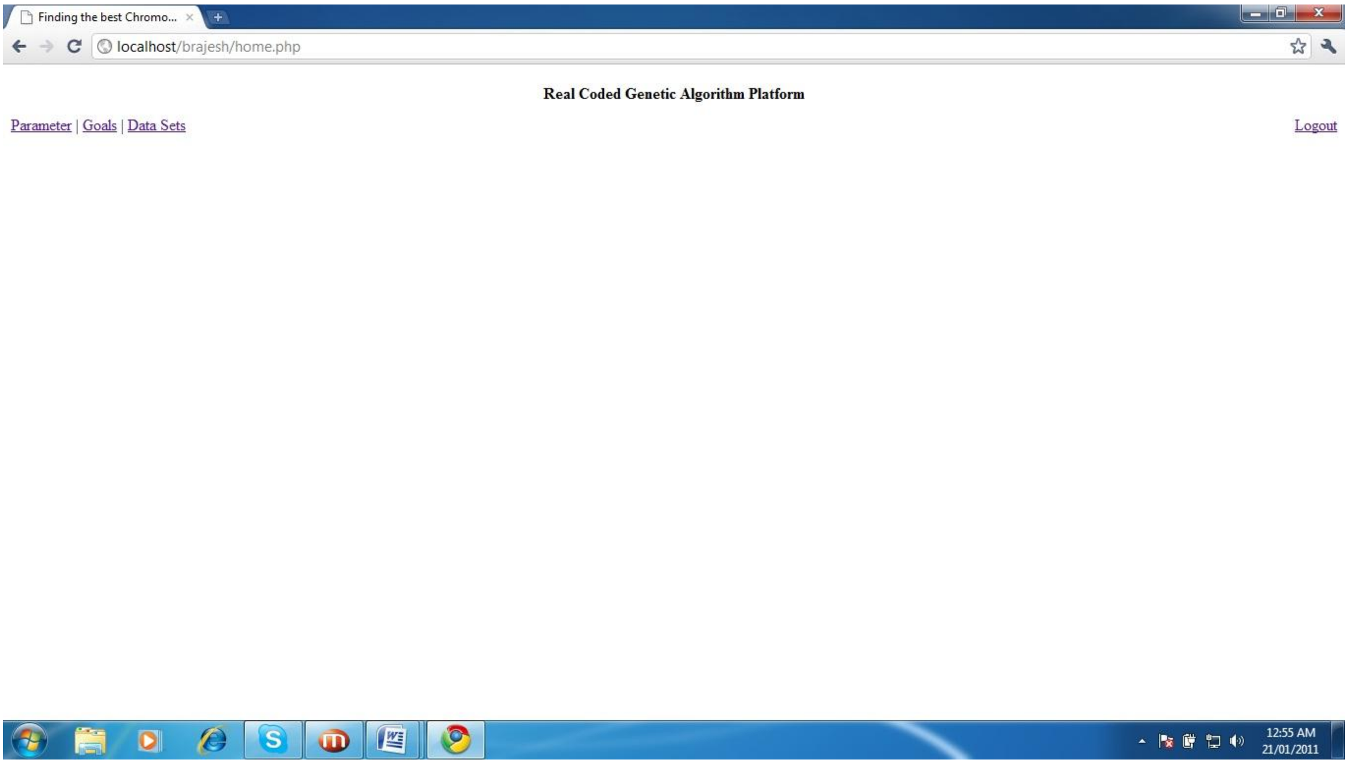
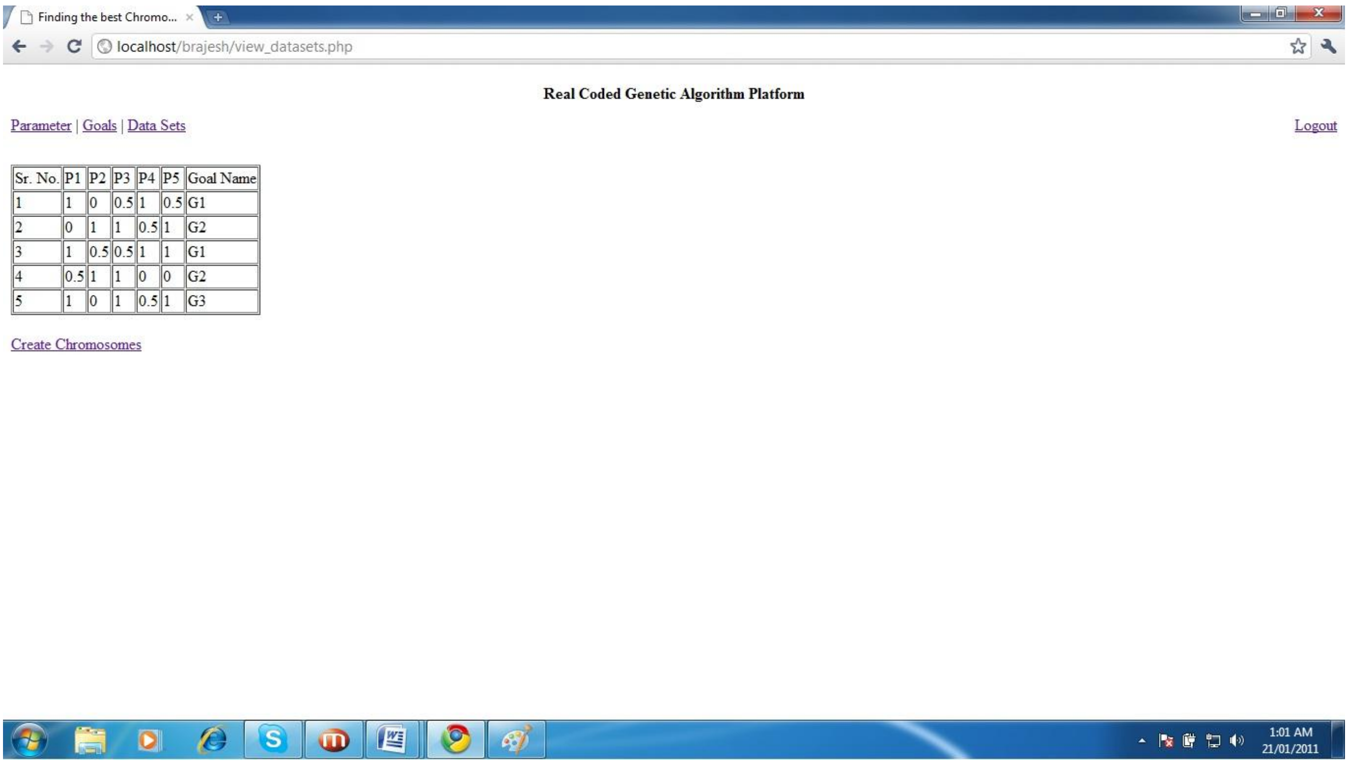
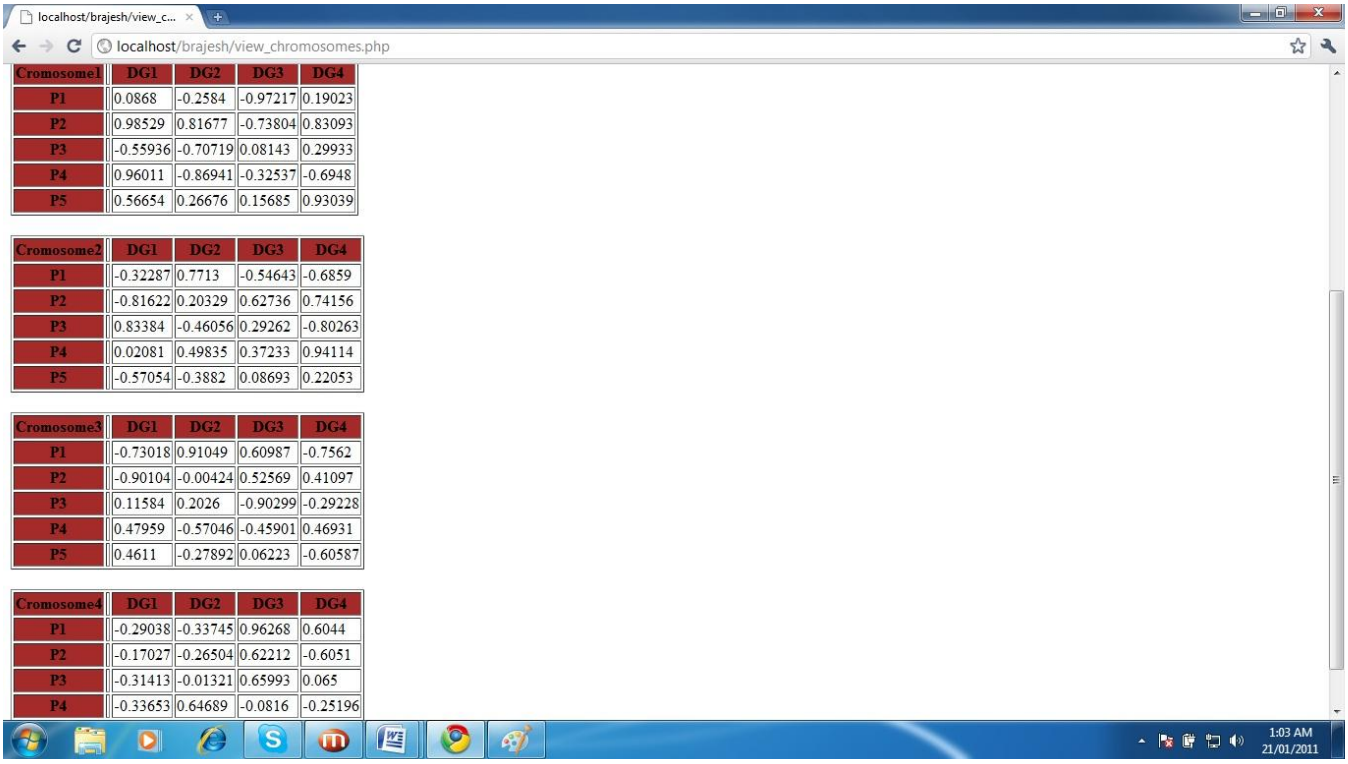


Fig 1: Genetic Algorithm Platform



Sr. No.	P1	P2	P3	P4	P5	Goal Name
1	1	0	0.5	1	0.5	G1
2	0	1	1	0.5	1	G2
3	1	0.5	0.5	1	1	G1
4	0.5	1	1	0	0	G2
5	1	0	1	0.5	1	G3

Fig2: Data Set with Goals and parameters



Cromosome1	DG1	DG2	DG3	DG4
P1	0.0868	-0.2584	-0.97217	0.19023
P2	0.98529	0.81677	-0.73804	0.83093
P3	-0.55936	-0.70719	0.08143	0.29933
P4	0.96011	-0.86941	-0.32537	-0.6948
P5	0.56654	0.26676	0.15685	0.93039

Cromosome2	DG1	DG2	DG3	DG4
P1	-0.32287	0.7713	-0.54643	-0.6859
P2	-0.81622	0.20329	0.62736	0.74156
P3	0.83384	-0.46056	0.29262	-0.80263
P4	0.02081	0.49835	0.37233	0.94114
P5	-0.57054	-0.3882	0.08693	0.22053

Cromosome3	DG1	DG2	DG3	DG4
P1	-0.73018	0.91049	0.60987	-0.7562
P2	-0.90104	-0.00424	0.52569	0.41097
P3	0.11584	0.2026	-0.90299	-0.29228
P4	0.47959	-0.57046	-0.45901	0.46931
P5	0.4611	-0.27892	0.06223	-0.60587

Cromosome4	DG1	DG2	DG3	DG4
P1	-0.29038	-0.33745	0.96268	0.6044
P2	-0.17027	-0.26504	0.62212	-0.6051
P3	-0.31413	-0.01321	0.65993	0.065
P4	-0.33653	0.64689	-0.0816	-0.25196

Fig3: Generation of Chromosomes

4. CONCLUSIONS

The genetic algorithm as mentioned above in section 2 is applied to the diagnosed database. The gene value of fittest chromosome indicates the importance of the parameter. Higher the value of gene, higher will be the importance of the parameter.The interpretation of the gene value is largely agreed by the experts.

5. ACKNOWLEDGMENTS

We wish to give our thanks to Mr. Somendra Shkla and Mr. Gaurav Yadav for discussions on this study

6. REFERENCES

[1] M.M.Raghuwanshi & O.G.Kakde, Survey on multiobjectibe evolutionary and real coded Genetic algorithms,www.complexity.org.au/conference/upload/raghuw01/raghuw01.pdf

[2] R.Setiono, Generating concise and accurate classification rules for breast breast cancer diagnosis, Artif. Intell. Med. 18 (2000) 205-219.

[3] Y.Hayashi, R.Setiono, K.Yoshida, A comparison between two neural network extraction techniques for the diagnosis of hepatobiliary disorders, Artif, Intell. Med. 20 (2000) 205-216

[4] http://en.wikipedia.org/wiki/Genetic_algorithm

[5] F. Herrera, M. Lozano & J.L. Verdegay, Tackling Real-Coded Genetic Algorithms: Operators and Tools for Behavioral Analysis, Artificial Intelligence Review 12: 265–319, 1998. 265

[6] Deb, K.M (2006): Optimization for Engineering Design, Algorithms and Examples, PHI, New Delhi