Volume 3, Issue 8, August 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Guidelines to Decide the Encoding Scheme Used For G.A.

Sumati jaggi
Skiet
Haryana, India

Ravi Bhushan
Skiet
Haryana, India

Davender Malhotra
D.A.V Girls Clg
Haryana, India

Abstract:- Genetic algorithm are optimization and robust searching technique that is used to find the best search result. It is an evolutionary approach that has been used in number of NP- hard optimization problems like travelling sales man problem, Dejong' s function, knapsack problem etc. Genetic Algorithm depends upon a various parameters as like type and probability of crossover and mutation operator, replacement operator etc. The main focus is how to represent the gene in a chromosome. Choosing the right scheme of encoding the gene is a crucial task. Encoding mainly depends on the type of problem. In this paper, various guidelines to decide the encoding scheme has been given. For this a number of encoding schemes like binary encoding, value encoding, permutation encoding, tree encoding etc. have been discussed and their applicability has been identified. A through survey of the encoding schemes was carried out and it is opined that the work performed will help the researchers in selecting a proper encoding scheme while solving any optimization problem using Genetic Algorithm.

Keywords:-chromosome, crossover , encoding, genetic algorithm.

I. INTRODUCTION

Genetic algorithms are inspired by darwin's theory of natural evolution. Genetic algorithm start with a set of solutions (represented by chromosomes) which called initial population [2]. Solutions from one population are taken and used to form a new population. The main operators of GA are selection, crossover, mutation and replacement. It is motivated by a hope that new population will be better than the old one. Selection of chromosome is inspired by the principal of survival of the fittest- better solutions have more chances to survive.
Encoding is a process of representing individual genes or one can say Encoding means to encode a solution of a problem into chromosome. Encoding is the way how genes are represented. Representation can also be termed as encoding [1]. A representation assigns genotypes to corresponding phenotypes. In genetic algorithms, genetic operators work on the level of genotype and evaluation of individuals is performed on the level of phenotype [4]. So, certain mapping or coding function between the phenotypic space and the genotypic space is needed. This can be done by designing the representation as close as possible to characteristics of phenotypic space. The mapping of the object variables to a string code is achieved through an encoding function and the mapping of a string code to its corresponding object variable is achieved through a decoding function. In simple words Encoding means representing genotype to phenotype[3].

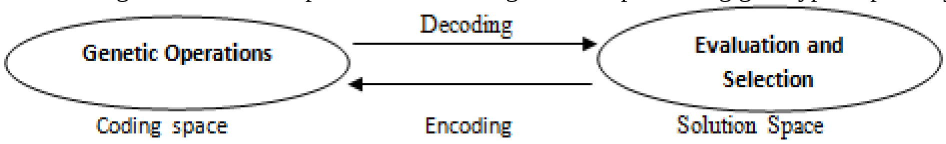

Fig 1: Encoding-Decoding method

The two principals to select an encoding schemes are [4]:-

1)Principle Of Meaningful Building Blocks:
A schema is a way of representing potential individuals that use bit string chromosome representation. Schemata are strings that consist of 0,1 and a wild card character. The principle of meaningful building blocks is directly motivated by the schema theorem. If schema is high-quality schemata are long or of high order, they are disrupted by crossover and mutation and they cannot be propagating properly.

2) Principle Of Minimal Alphabets:-
The alphabet of the encoding scheme should be as small as possible while still allowing a natural representation of solutions. It states that the user should select the smallest alphabet that permits an expression of the problem so that the number of exploitable schemas is maximized.
Encoding scheme can be classified on the basis of:-

- Encoding scheme whose fitness depends upon value of allele only.
- Encoding scheme whose fitness depends on the value and order both.

The main issue is how to represent the genes in a chromosome. Choosing the right scheme of encoding the genes is a crucial task. Encoding mainly depends on the type of problem. On the basis of problems, encoding categorize into various types as binary, permutation, value encoding etc.

II. TYPES OF ENCODING

2.1)<u>Binary Encoding scheme:</u> -

It is the simplest way to represent chromosomes. Binary Encoding means encoding using binary strings. A Binary String is a sequence of characters from { 0, 1 }.Thus in Genetic Algorithm

Solutions are represented in Binary Encoding as string of 0s and 1s. This encoding type mainly used in problems like knapsack problem. This is one of the popular encoding scheme as earlier GA work used this type of encoding.

Table 1: Binary encoding

Chromosome 1 →	00110010	Fitness= 50
Chromosome 2 →	01010101	Fitness= 84

<u>Operators For Binary Encoding Procedure</u> :- One can use crossover and mutation as straight forward way in binary encoding scheme . Crossover operator is a genetic operator that combines two chromosomes to produce a new chromosome (offspring). One can apply single point, two point crossover and uniform crossover in binary encoding scheme. The main idea behind crossover is that the new chromosome (solution) may be better than both of the parents (solution) & it takes the best characteristics from each of the parents**.**

<u>Single Point Crossover:-</u> In this one cross point is generated between 1 and chromosome size. Then genes before cross site are copied from 1st parent and genes after cross site from 2nd parent are copied to 1st child same way is applied to the second child. In below figure the crossover point is 2. As shown in below figure that the first offspring is formed by taking the values before the crossover point from the first parent i.e. 0 1 and rest part is formed by taking the values after the crossover point from the second parent i.e. 1 0 0 1 1 and in case of second offspring formation taking before part from second and after part from first, as shown:-

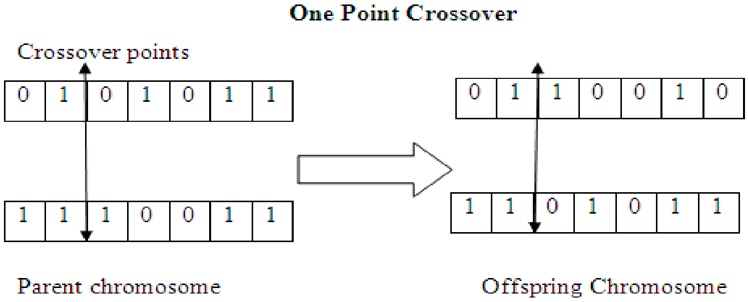

Fig 2: single point crossover

In figure 2, the arrows indicate the crossover points. Thus, the contents between these points are exchanged between the parents to produce new children for mating in the next generation.

<u>Two Point Crossover:-</u> Many different crossover algorithms have been devised, often involving more than one cut point, unlike Single point Crossover. It should be noted that adding further crossover points reduces the performance of the GA. The problem with adding additional crossover points is that building blocks are more likely to be disrupted. However, an advantage of having more crossover points is that the problem space may be searched more thoroughly. In two-point crossover (TPC), two crossover points are chosen and the contents between these points are exchanged between two mated parents. In figure, the arrows indicate the crossover points.

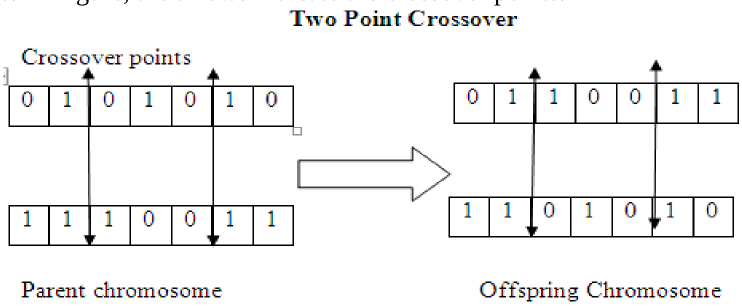

Fig 3: Two point crossover

<u>N-point Crossover:-</u>This crossover is generalization of single point crossover. N random points are selected. Split the chromosome along these points. A new child is created by gluing parts alternating between parents as shown in Table 2. As this is generalization of single point crossover in the sense that it is highly exploratory and when population size is small then this crossover is recommend.

Table 2: N point Crossover

Parent 1	000	000	000	000
Parent 2	111	111	111	111

Offspring 1	000111000111
Offspring 2	111000111000

Coding in matlab for representation of binary encoding schemes:-

```
x=input('enter number of individual');
y=input('enter size of individual');
geno=round(rand(x,y));
```

x is no of individual , n is size of individual & geno is x*y matrix in binary encoding scheme fitness depends on value and order: f(v,o).

Mutation Operator:-

In mutation, one bit is selected & selected bit is inverted. Altering one or more gene value in a chromosome is called mutation and the Mutation operator changes the values of genes. Let mutation site is 5**.**

Table 3: Mutation operator

Chromosome	**11001101**
Offspring	**11000101**

2.2)Permutation Encoding:-

Permutation encoding can be used in ordering problems, because binary encoding scheme cannot be able to solve ordering problems such as traveling salesman problem or task ordering problem. In permutation encoding, every chromosome is a string of numbers that represent a position in a sequence. In TSP there are N cities starting from first city and tries to shorten the distance. In other words, find a minimal Hamiltonian circuit in complete graph of N nodes. To represent order of cities we cannot use binary encoding. Chromosome is represented as sequence of integers, where each integer represents a different city and order represents the time at which city is visited. So, in this case we cannot use binary encoding scheme.

Operators in permutation encoding schemes:- The crossover and mutation operators used in permutation encoding scheme are not direct forward as in binary encoding schemes. Crossover operators which are used for permutation encoding are partially Matched Crossover, Order Crossover. Crossover operators formulated as:-

- PMX
- OX

Partially Matched Crossover (PMX):- The PMX technique builds offspring by choosing a subsequence of a tour from one parent and preserving the order and positions of as many cities as possible. In it, two strings are aligned, and two crossover points are selected uniformly at random along the length of the strings, and that points acts as boundaries for the swap operation.. In below Table Two crossover points were selected at random, and PMX proceeds by position wise exchanges. In-between the crossover points the genes get exchanged i.e., the 3 and the 2, the 6 and the 7, the 5 and the 9 exchange places. This is by mapping parent B to parent A. Now mapping parent A to parent B, the 7 and the 6, the 9 and the 5, the 2 and the 3 exchange places. Thus after PMX, the offspring produced as follows. Consider two strings:

Table 4: PMX

Parent A	4 8 7	3 6 5	1 10 9 2
Parent B	3 1 4	2 7 9	10 8 6 5

Child A	4 8 6	2 7 9	1 10 5 3
Child B	3 1 4	3 6 5	10 8 7 9

where each offspring contains ordering information partially determined by each of its parents. PMX can be applied to problems with permutation representation.

Order Crossover (OX):- The order crossover begins in a manner similar to PMX. But instead of using point by point exchanges as PMX does, order crossover applies sliding motion. One important point about PMX and OX is that PMX preserves absolute order while OX preserves relative order. In order to create an offspring the String between the two cut points in first parent is first copied to the offspring. Then the remaining positions are filled by considering the sequence in the second parent, starting after the second cut point(when the end of chromosome is reached, the sequence continues at position 1) . this is shows in the below table as:-

Table 5: OX

Parent A	4 8 7	3 6 5	1 10 9 2
Parent B	3 1 4	2 7 9	10 8 6 5
Child A	3 6 5	2 7 9	1 10 4 8
Child B	2 7 9	3 6 5	1 0 8 1 4

From the examples, it can be noted that PMX tends to respect absolute positions While OX tends to respect relative positions.

<u>Mutation operator</u>:- For permutation encoding swap mutation, inversion mutation, scramble mutation etc. are used.

<u>Swap mutation:</u>- [1] As swap meaning changing the values of selected element. Same in swap mutation means change the elements in selected position such as:-

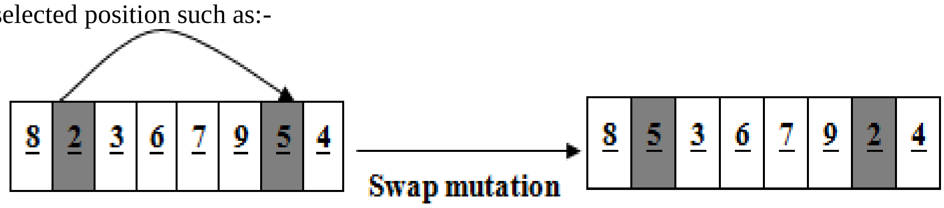

Fig 4: swap mutation

<u>Inversion mutation:-</u>The order of elements between the selected locations is inverted in this mutation.

<u>Value Encoding:-</u> In some problems where some more complicated values such as real numbers are used use of binary encoding is difficult, so Direct value encoding can be used for these type of problems . In this encoding, every chromosome is a sequence of some values. Values can be anything related with the problem, such as (real) numbers, chars or any objects.

Table 6: value encoding

Chromosome A	1.2324 5.3243 0.4556 2.3293 2.4545
Chromosome B	ABDJEIFJDHDIERJFDLDFLFEGT
Chromosome C	(back), (back), (right), (forward), (left)

<u>Coding in matlab for representation of value encoding(real value) schemes:-</u>

x= input('enter number of individuals');
y=input('enter size of individuals');
geno= Rands(x,y);

Here x is number of individuals & y is size of individual.

Operator in value encoding scheme:-

Arithmetic crossover:-

This operator combine two chromosome & produce 2 offspring by formula.

Offspring1= a * parent 1+(1-a) * parent 2

Offspring 2=(1-a) * parent 1+ a * parent 2

<u>Tree Encoding</u>:- Tree encoding is two dimensional encoding scheme. The special use of tree encoding is that it is used for evolving programs or expressions, i.e. for genetic programming. Genetic programming is specialized part of GA. In the tree encoding every chromosome is a tree of some objects, such as all leafs are values or variables and internal nodes represent functions

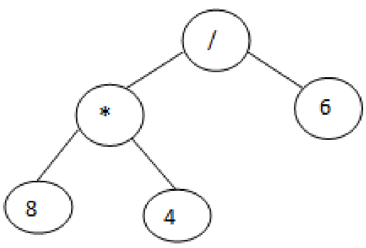

Fig:5- Tree encoding

Crossover operator used in operators in tree encoding, shown in below diagrams as:

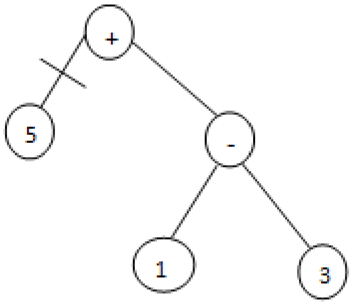

Fig:6-a

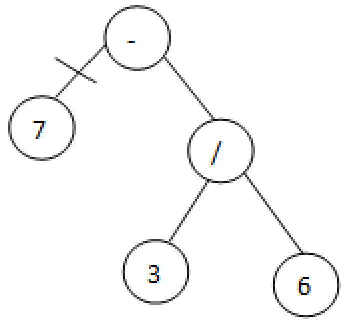

Fig:6-b

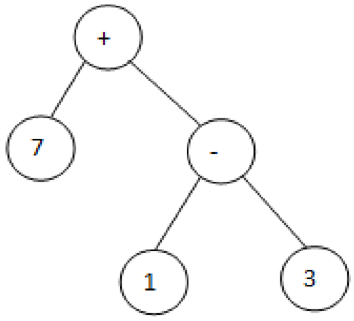

Fig: 6- c

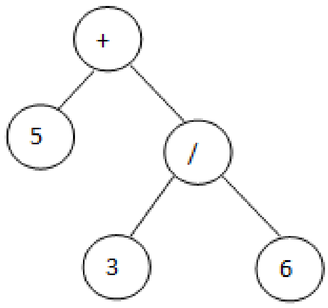

Fig: 6- d

Fig 6: crossover operator in tree encoding

<u>Octal Encoding</u>:- Octal Encoding is same as Binary Encoding. As octal sequence is from(0-7) So, String in Octal Encoding is made of octal numbers (0-7).

Table 8: Octal encoding

2 3 7 1 4 0 5 6	← Chromosome

<u>Hexadecimal Encoding</u>:-Same like octal encoding Hexadecimal Encoding is also same as Binary Encoding. Ans String in Hexadecimal Encoding is made of hexadecimal numbers (0-9 , A-F).

Table 9: Hexadecimal encoding

3 D 1 F 7 A 4 1	← Chromosome

<u>Floating Point Encoding</u>:- In Floating Point Encoding gene has one of the three elements. Three elements are :-

1. Parameter Identification Number
2. Mantissa or Exponent
3. Value

III. CONCLUSION

After study of existing encoding schemes, it has been found that encoding scheme used for representation of chromosome and this representation should follow two principals, that increases the survival rate of individuals i.e. principal of meaningful building blocks and principle of minimal alphabets. In some encoding schemes, as in various types the fitness depends on order of genes while in other it depends on value. Different encoding schemes require different crossover and mutation operators. There is possibility that an operator applicable in one may or may not be applicable in others. Encoding schemes applicability has lot to do with the problem specification.

REFERENCES

[1] D.E.Goldeberg, Genetic algorithm search, optimization and machine learning. Addition Wesley Longman, Inc, ISBN0-201-15767-5, 1989.

[2] S. N. Shivanandam, S.N.Deepa, Introduction to Genetic Algorithms, ISBN 978-3-540-73189-4 Springer Berlin Heidelberg New York.

[3] Rakesh Kumar, Jyotishree, Novel Encoding Scheme in Genetic Algorithms for Better Fitness

[4] Anit kumar, Encoding Schemes In Genetic Algorithm, ISSN: 2278-6244, Vol. 2 | No. 3 | March 2013

[5] N. cahiyarataiia, A.M.S Zalzala ,Recent Developments In Evolutionary And Genetic Algorithms Theory And ApplicationsGenetic algoruithm in engering systems:Innovations and applications,2-4 septemeber 1997,Confrence Publication No. 446,IEEE,1997

