International Journal of Research and Reviews in Computer Science (IJRRCS) Vol. 2, No. 1, March 2011

Representing System by Gene Encoding

Ahmad Yusairi Bani Hashim

Department of Robotics & Automation, Faculty of Manufacturing Engineering,
Universiti Teknikal Malaysia Melaka,
Hang Tuah Jaya, Durian Tunggal, 76100 Melaka, Malaysia

***Abstract*:** The chromosomes that belong to the parents store genetic information. The child inherits the information through uniquely selective procedures. Using the concept, this study proposes an approach to system representation using the process found in genetic information inheritance. A generic process flow to building the representation is shown in the form of ladder-like flow chart that has the starting point at the bottom. This forms the basis for gene encoding. Following this process flow, one can expect to represent a system which requires solutions from selective procedures. Therefore, the codes written may vary depending on how one models and interprets the representation.

***Keywords*:** Genetic information, chromosome, selective procedure**.**

1. Introduction

The genetic information that comes from the chromosomes form strings that define a particular solution. If 110011 represent a chromosome, then a population of these chromosomes corresponding to a number of individual solutions to the problem is created.

There will be possibly some good solutions given by a population of chromosomes. The next process is to measure the fitness of each population of chromosomes. In addition, the selected fitness function is specific to the particular application, but generally it has a positive value when the solution is good, whereas small value represents a bad solution.

For example, if parent number 1 has chromosome 110011 and parent number 2 has chromosome 101010, then by crossing over some of the genetic materials there exist offspring number 1 chromosome 110010 and offspring number 2 chromosomes 101011.

This crossing over is called the single-point crossover. The process in crossing over actually is the breeding of a new population. While a new population is being created, mutation is allowed. A mutation is defined by the mutation rate. For example a chromosome of a new population is 1110010101010 can mutate by a new chromosome 111001**1**101010. The bold digit changed from 0 to 1 showed mutation has occurred.

This concept is generally accepted in genetic algorithm. In this paper, however, we introduce a concept of genetic information inheritance. It should be suitable for representing a system which requires solutions from selective procedures.

2. Gene Encoding

The diagram in Figure 1 explains the steps in realizing the gene encoding (GE). It has eight significant processes. The steps begin with creating population from data obtained follows with the number of digits set for each chromosomes, defines gender and set union rules.

Further, the number of offspring obtainable from a family and the rules for transmission of genes from one generation to another are set. There could be possibilities to adopt offspring in case the family is not fit to produce one.

The rules for the most qualified offspring is established and how they relate to their ancestors is investigated. Lastly, it is to see if the selected ancestors become the sample that possess the best chromosomes. The follwing section will explain the steps depicted in Fig. 1.

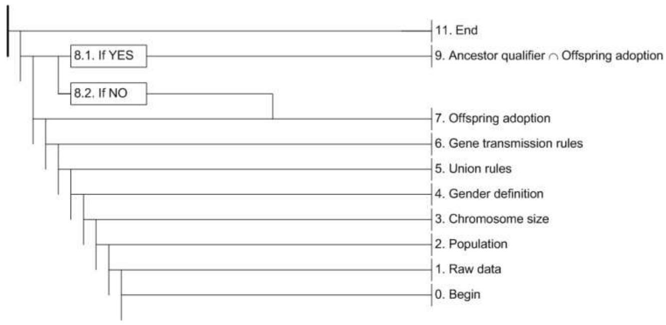

Figure 1. The steps taken in realizing the GE. The sequence is read bottom-up. From the initial sequence, the sequence begins with raw data assembly and follows by the determination of data population.

3. Representation

A population is created based on the type of input data. Every population is unique. There will be no one population that shares a type of input data. If there is more than one input types then the population of the corresponding inputs may be created.

In short, there will be a number of different sets of population. For example, when in a system there are two types of input data: angle, r and position, p. In this case, there are two different populations of r and p respectively. Input data can be described as a function such as $f(x)$, where x is the input data.

A population is then converted into strings or chromosomes. The method for converting the populations to chromosomes can be done in many ways. For example, the chromosomes can be obtained by defined encoding schemes where a permutation of all genes becomes a genetic chromosome [1]. A series of operation is used to represent a chromosome such that in the operation that yields to scheduling of job-shop and flow-shop in a manufacturing system [2]. In addition, chromosome can be represented as genetic string by applying objective functions of *x* and *y* where through parameter coding [3]. One-half of the genetic string is resulted from coding of *x* and another half comes from coding of *y*. The length of a genetic string is to be defined and optionally the size of a chromosome.

Researches in genetic algorithm typically related to engineering application that define gender for particular chromosomes in a population have not been found elsewhere except in [4] where the work claims that all chromosomes are of both sexes. Clownfish, for example, is naturally able to change gender.

Obviously, a population that is created from input data is a living entity. It is of our full control to assume the kind of organic structure that is desired. Either the chromosome is of male or female or both sexes depend largely on the end results expected.

To have offspring there must be marriages. Unless there is a certain orientation of chromosomes that has been defined to represent a gender then the process for blending can take place—the union. In [3], the cross-over process is applied to produce offspring. There are two offspring resulted from the crossovers. However, any information containing different types of gene would generate illegal offspring [2].

The parents are chosen randomly and by partial cross-over two offspring are produced. Could there be twins? One can define that for each union a certain number of offspring are allowed.

If a person possesses two of each chromosome, and two of each genes type on each of a chromosome, he or she can have two genes for eye color [5]. When both genes of a given pair of chromosomes are identical, the person is said to be homozygous for that gene.

On the other hand, when the two genes are different, the person is heterozygous. There are genes called dominant genes and recessive genes. Dominant genes are able to exert their effects on development even in a person who is heterozygous for that gene. Recessive genes possess by those of homozygous' only. A heterozygous' will show the effects of the dominant gene but still pass the recessive genes to a son or a daughter.

It is necessary to have rules for transmission of genes from one generation to another. In some applications there might exist the requirement for producing the heterozygous offspring.

There is likely to occur a situation where offspring are not produced but the new population must expand. If such a case occurs then offspring may be adopted. The orientation of genetic strings of an adopted offspring must have a strong resemblance to its original parents. The genetic conditions in adoption process are, however, of the designer's full control.

In a natural selection, in Charles Darwin's theory of evolution, is that individuals with certain genetically controlled characteristic reproduce more successfully than others. The artificial selection found in animal and plant breeders is used for generating new strains. Moreover, a person who has five healthy children before dying at age of thirty is evolutionary successful [5]. For the new population created, the qualified ancestors or the original parents can be said to be evolutionary successful because their unions have successfully produced offspring with unique genetic strings patterns. Consequently, the qualified ancestors are said to have the most eligible chromosomes.

4. Model

We assume that every test would result in unique codes. A test is called candidate, C . The sum of C becomes the population of a culture. Equation (1) exhibits how G_C is represented. The symbol $\cap$ does not portray intersection or union as in theory of sets. It describes the process of matching C or the marriage, M . Only if a couple that has the same genetic codes, then they can have a child, ch .

Equation (2) explains any union of C can only result a child. It is agreed that any unions that produce a ch reflects to the best combination of ancestors. The ancestors are those who start the population. Any unions that do not produce a ch are considered as the unfit ancestors, hence are abandoned. This is summarized in Figure 2 where three candidates X , Y , and Z that are producing their successors.

$$G_C = \left\{ g_i \in \text{bit} \;\middle|\; g_i = \begin{cases} 1\text{xxxx} & \text{if exist} \\ 0\text{xxxx} & \text{otherwise} \end{cases} \right\} \tag{1}$$

$$M_k = \left\{ \left\{ C_i \cap C_j \right\} \in \text{bit} \;\middle|\; m_k = \begin{cases} 1 & \text{qualified ancestor} \\ 0 & \text{unfit ancestor} \end{cases} \right\} \tag{2}$$

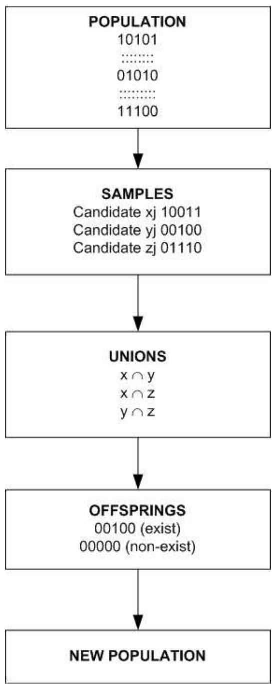

Figure 2. While Figure 1 is the general flow for GE this flow chart, however, depicts of the processes executed using eqn. (1) and eqn. (2).

5. Discussion

The GE shown in Fig. 1 is a general representation and is open for modification. The common approach to visualize this diagram is by a typical flow chart seen in Fig.3. We avoid this type of chart because it is exact and the blocks consume space. In addition, it stores information by the shape of the blocks, which we think is specific and fixed.

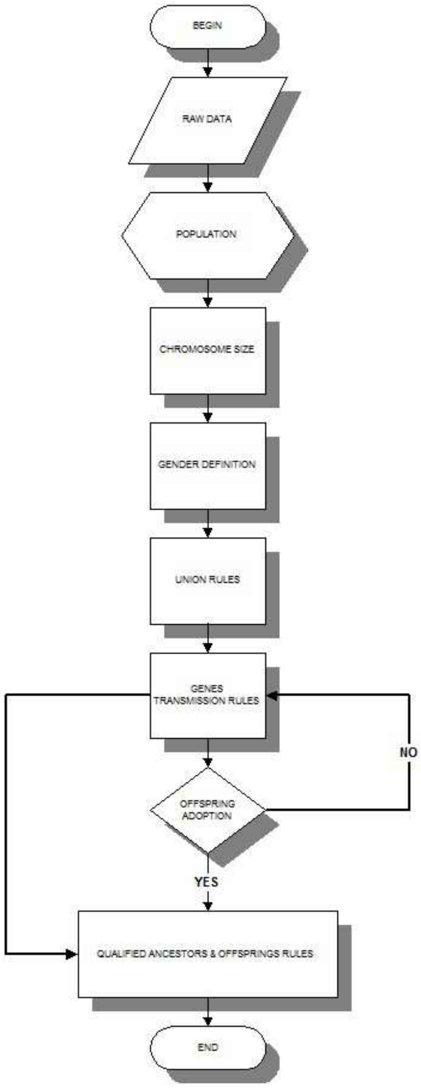

Figure 3. Another way of describing the GE as opposed to Fig.1.

Our proposed GE is straight forward but is flexible. The flexibility is found in the gender definition, union rules, gene transmission rules, and offspring adoption. In gene expression programming (GEP), however, the constrains of the head-tail mechanism contribute to the legality of chromosome [6]. In GE, legality of chromosome is based on the procedures defined in the union rules and the adoption of offspring.

On the other hand, cluster analysis is a major method to study gene function and gene regulation information for there is lack of prior knowledge for gene data. Many clustering methods existed at present usually need manual operation or pre-determined parameters, which are difficult for gene data [7]. In GE, however, there is no clustering of data except during the process of population determination. This operation can be set manual or automatic depending on the type of applications. If the application is distinct and occasional then the operation is manual. Otherwise, it can be set automatic if the application is routine and repeating.

Normally, in GEP, each chromosome is expressed and evaluated in on the Expression Tree (ET). The ET-based expression and evaluation are computationally expensive and the intelligibility of the chromosome is low [8]. In GE, the chromosome is not evaluated on the ET. It exists simply though procedures in population determination and after the union rules have been executed. In fact, a chromosome can be added if there is a need for offspring adoption. The inheritance is based on indefinite cross-over.

6. Conclusion

The method has its strength in analyzing sets of data and gives solutions from selective procedures. This procedures are predefined by the designer. It is of the designer's total control on how algorithms are arranged and manipulated. Therefore, the steps in building the codes discussed in this paper can be made as mere guidelines for specific software design.

Acknowledgement

Thanks to Universiti Teknikal Malaysia Melaka for providing sufficient resources and facilities for research activities.

References

[1] W. Wang and P. Brunn, P., "An effective genetic algorithm for job shop scheduling," Proc. Institution of Mechanical Engineers, 214B, pp. 293-300, 2000.

[2] L.P. Khoo, S.G. Lee and Yin, X.F., "GA-based decision support systems in production scheduling", The Int. J. of Adv. Manufacturing Technology, 8, 131-138, 2000.

[3] H.S. Ismail and Hon K.K.B. "The nesting of two-dimensional shapes using genetic algorithms," Proc. Institution of Mechanical Engineers, 209, pp. 115-124, 1995.

[4] A.Y. Bani Hashim, "Development of artificial intelligent techniques for manipulator position control," Masters Thesis, Universiti Putra Malaysia, 2003.

[5] J.W. Kalat, "Introduction to psychology," Third edition, Brooks/Cole Publishing, California, 1993.

[6] J. Zhang, Z. Wu, Z. Wang, J. Guo, Z. Huang, "Unconstrained gene expression programming," Proc. of the IEEE Congress on Evolutionary Computation, pp. 2043-2048, 2009.

[7] G. Liu, X. Jiang, L. Wen, "A clustering system for gene expression data based upon genetic programming and the HS-Model," Proc. of the third International Conference on Computational Science and Optimization (CSO), pp. 238-241, 2010.

[8] Y. Chen, C.J. Tang, R. Li, M.F. Zhu, C. Li, J. Zuo, "Reduced-GEP: Improving gene expression programming by gene reduction," Proc. of the 2nd International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), pp. 176-179, 2010.