EFFECT OF CROSSOVER AND MUTATION OPERATORS ON HAMMING DISTANCE OF THE OFF-SPRINGS IN GENETIC ALGORITHM

[1]FARHAD NADI, [2]AHAMAD TJUDIN KHADER

[1]School of Computer Sciences, Universiti Sains Malaysia, Penang, Malaysia
[2]School of Computer Sciences, Universiti Sains Malaysia, Penang, Malaysia
E-mail: [1]nadi@cs.usm.my, [2]tajudin@cs.usm.my

ABSTRACT

Crossover and mutation operators are the only operators that change the values of the loci in genetic algorithm. New solution would be generated using this two operators and with the aid of selection operators the search will be directed towards the promising solutions. The resulted solutions of these operators are highly dependent on their parameters. Crossover operators take two parent solutions as input and create two new offspring solutions which will be resulted from mixing the parent solutions. Mutation operator however, is a unary operator that takes a solution and randomly change the values of some of the loci. The resulted offspring would have different distances from their original parents. In this research the distance of the parent solution and their offspring will be studied. It has been shown that mutation operator could reduce the distance between parent and the off-springs.

Keywords: *Genetic Algorithms, Parameter Control, Parametric uniform crossover, Exploration, Exploitation*

1. INTRODUCTION

Recombination operator plays an important role in Genetic Algorithms (GAs). In early days of GAs, n-point crossover operators with small value for n were recommended based on theoretical and empirical foundations [1, 2, 3]. However, some later researches have shown that larger values for the number of crossing points (n) could also be beneficial in some circumstances [4, 5].

Random re-sampling of the individuals would change distribution of the solutions in the search space. A uniformly distributed search space, or a diverse population, would facilitate exploration of the search space. As the search progress the population converges toward a solution in the search space. Resampling of a converged population has the cost of losing the information of the existing good solutions. In other words, re-sampling of the population would cause destruction of the existing good (partial) solutions.

It has been shown that different crossover operators have different destructive effects [3]. Mutation operator has some destructive effects as well; however its effect is negligible due to its low rate, most of the time. Parametric Uniform Crossover (PUC) is a general form of uniform crossover operator on which the exchange probability of the loci (p_0) could be controlled. Unlike the uniform crossover where its destructive effect is fixed, the destructive effect of the PUC could be controlled by the value of the p_0 [3].

Existence of a proper balance between exploration and exploitation of the search [1] is of importance in searching for high performance quality solutions. The amount of destruction in hyperplanes could be seen as exploration of the search space, which will be done at the expense of exploitation. Thus, varying the destruction level of the hyperplanes changes the balance between exploration and exploitation [3].

Uniform crossover has the most effect on destructing building blocks, in comparison to the n-point crossover operators. The destructive effect of the other operators is dependent to the hyperplane defining length. However, the uniform crossover has the advantage that its destructive effect is independent of the hyperplane defining length [3].

Besides the uniform crossover, the mutation operator has also effect on resampling of the

solutions. The mutation operator is usually expected to introduce small changes to the solutions. However, there has been some researches that have shown higher mutation rates could be beneficial as well [6].

2. PARAMETERISED UNIFORM CROSSOVER

Algorithm 1 shows a pseudo code of the uniform crossover assuming that $p_0 = 0.5$ (see line 4 in Algorithm 1) is a constant. This Algorithm would represent the pseudo code of the PUC by considering p_0 as a parameter of the operator. The uniform crossover operator exchanges the alleles of two loci with probability of 0.5. It would cause exchange of $l/2$ of the alleles on average [3]. The exchange of the alleles of two parents would result in two new off-springs that are differing from their parents in genotypic space and consequently, in phenotypic space. Hamming distance could be used to measure the genotypic distance of the off-springs from their parents. Using uniform crossover, there always is a relationship between the parents and the resulted off-springs. The farther the parents are from each other in hamming distance, the longer is the distance between them and vice versa (see Figure 1a).

The hamming distance between two given individuals could be calculated using,

$$h(x, y) = \sum_{i=1}^{l} x_i + y_i - 2x_i y_i$$
$$= \sum_{i=1}^{l} x_i \oplus y_i \tag{1}$$

Where l is the length of the individuals, and $\oplus$ represents the binary XOR operator.

Algorithm 1: General pseudo code for Uniform crossover operator. If p0 is fixed to 0:5 the operators is equal to the traditional Uniform crossover, while if the value of with p0 be treated as a parameter, the operator will be the parametric Uniform crossover.

Data: I_1, I_2, P_c, p_0
// I_1, I_2 represent parent solutions
Result: I_1', I_2'
// I_1', I_2' represent off-spring solutions

```
1   if  U(0,1) < P_c
    then
2   |   l = I_1.length
3   |   For i = 1 to l do
4   |       if
5   |       |   I'_2[i] = I_1[i]
6   |       |   I'_1[i] = I_2[i]
7   |       else
8   |       |   I'_1[i] = I_1[i]
9   |       |   I'_2[i] = I_2[i]
10  |       end
11  |   end
12  end
```

The distance between two given parents (δ) would be equal to the distance between their resulted off-springs regardless of the crossover operator. This equality holds for all types of crossover operators as in all of the crossover operators, the alleles would only be swapped between the loci of two individuals.

Similar parents would have smaller Hamming distance and crossover on them would also result in solutions that are near (homogeneous) to each other. As it has been shown, the uniform crossover would always create two off-springs at the midpoint in the hamming space. However, using the PUC and thanks to the ability for changing the swap rate (p_0), it would be possible to mix the individuals in such a way that their hamming distances are not equal with their parents. In other words, PUC is capable of mixing the genetic materials in a biased way (see Figure 1b). As like in the uniform crossover, the distance between the parents, δ, is equal to the distance between the off-springs for the PUC operator as well. If the distance of one of the off-springs with one of the parents is j, then the distance of that off-spring with the other parent would be k under the condition that $j + k = \delta$. In the cases where $j \approx k \approx \delta/2$, the PUC would perform like uniform crossover.

In the following, it will be assumed that two parent individuals, A and B are negation of each other. This will be without the loss of generality, as the hamming distance for the loci with similar allele is   zero.
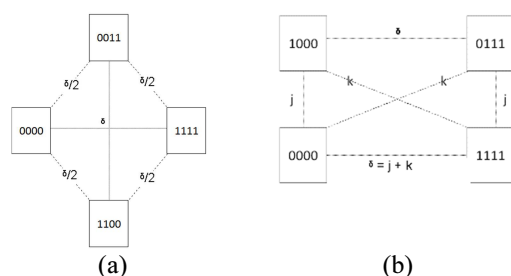
Figure 1: Crossover Operator Mixes The Genetic Material Of Each Parent To Create New Off-Springs.(A) Uniform Crossover Will Create Two Off-Springs With The Same Distance From Each Of Parents.(B) PUC Is Capable Of Creating Off-Springs Which Are Biased Towards The Parents

Let's assume that, δ is the number of the alleles which are exchanged by the application of uniform crossover. Regarding to the commutative property of the Hamming distance, the alleles in the individuals could be reordered in such a way that all of the swapped alleles appear in one end of the individuals. This way, the parent individuals (A, B), and off-springs (A', B') could be reordered into two parts namely fix and swapped parts, denoted by **F** and **S** subscripts, respectively,

$$A = A_F + A_S, B = B_F + B_S \Rightarrow A' = A_F + B_S, B' = B_F + A_S.$$

By supposing,

$$h(A, B) = h(A', B') = |A| = l = \delta,$$

It could be concluded that,

$$
\begin{aligned}
h(A, A') &= h(A_F + A_S, A_F, + B_S) \\
&= h(B_F + B_S, B_F, + A_S) \quad (\text{2}) \\
&\approx |A_S| \\
&\approx |B_S|
\end{aligned}
$$

This way, the hamming distance between the A, B' and B, A' would be,

$$
\begin{aligned}
h(A, B') &= h(B, A') \\
&\approx l - |A_S| \approx l - |B_S|.
\end{aligned}
$$

As mentioned earlier, the percentage of the exchange of the alleles in uniform crossover is 50% thus $|A_S| \approx |B_S| \approx l/2$ in Eq. (2). To this end, the distance between the parents and both of the off-springs will be the same ($\approx \delta/2$) as depicted in Figure 1a ($\delta \approx l/2$). In the other hand, for the PUC the value of p_0 determine the percentage of the exchange (mix) between two parents. Therefore the distance between the parents and off-springs would be non-symmetric as shown in Figure 1b.

It has to be noted that the hamming distance of two randomly generated individuals is $l/2$ on average. Nevertheless, this value has been assumed to be l in the above equalities. This way, the average amount of exchanges in loci would be $\frac{l/2}{2} = \frac{l}{4}$ or 25%, as from total of $l/2$ exchanges half of them are similar alleles and swapping them would not make any difference.

To this end, it could be inferred that the resulted off-springs would be far from each other if their corresponding parents are far from each other and vice versa. As the search starts with a random initialisation of the population, the similarity of the individuals is low (their hamming distance is high). Therefore, application of crossover operators could result in non-homogeneous off-springs.

Selection pressure would gradually increase the homogeneity of the population. As a result, the hamming distance of the selected parents will decrease and likewise will be the distance of the resulted off-springs. Accordingly, it could be inferred that the crossover explores the search space in the beginning of the search. However, the exploratory role of the crossover changes into exploitation of the search space as time passes as fewer exchanges would occur as the search progresses.

3. EFFECT OF INVERSE MUTATION ON HAMMING DISTANCE

There are several different mutation operators that could be used for binary GAs. In this research inverse mutation will be considered for analysis. This operator would simply reverse the allele of each loci if a drawn random number is less than mutation rate p_m. In the following effect of mutation on hamming distance of the mutated off-springs, off-springs, and their parents with each other will be studied.

1.1 Uniform Crossover

Assuming that $h(p_1, p_2) \approx \frac{l}{2}$, the following equalities will hold for uniform crossover. As shown in Figure 2 the distances between (p_1, c_1), (c_1, c_1'), and (p_1, c_1') would be as follows.

$$h(p_1, c_1) \approx \frac{l}{2} = \frac{\delta}{2}$$
$$h(c_1, c_1') \approx l.p_m \qquad (3)$$
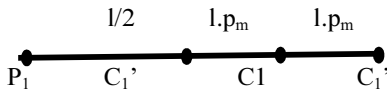$$h(p_1, c_1') \approx \frac{l}{2} \pm l.p_m$$

Figure 2: The Distance Of Parent With The Child Solution In Uniform Crossover

Where p_1 is a given parent, c_1 is its off-spring over corssover, and c_1' is the mutated off-spring. As it has been shown earlier, a given individual $A = A_F + A_S$, would be changed into $A' = A_F + B_S$ after crossing over. As it has been discussed earlier under uniform crossover $|A_F| \approx |A_S| \approx \frac{l}{2}$.

By applying mutation operator $l.p_m$ alleles would change. As any allele could be mutated independent of other alleles, it could be expected that $|A_F|.p_m$ loci will be mutated in the first part of the offspring and $|B_s|.p_m$ mutations would happen in the second (swapped) part of the offspring. Changes in the first part of the offspring would increase the hamming distance of the mutated offspring from the parent solution while mutations in the second part of the offspring would reduce this distance.

1.2 Parameterised Uniform Crossover

Assuming the Hamming distance between two parents $h(P_1, P_2) \approx l \approx \delta$, the following equalities will hold for uniform crossover.
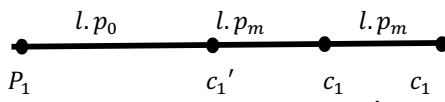

Figure 3: The Distance Of Parent With The Child Solution In Parameterised Uniform Crossover

According to Figure 3 the following equalities hold.

$$h(p_1, c_1) \approx l.p_0$$
$$h(c_1, c_1') \approx l.p_m \qquad (4)$$
$$h(p_1, c_1') \approx l.p_0 \pm l.p_m$$

As mentioned earlier, over the PUC $p_0 \leq 0.5$. This way, the $|A_F| \approx l - |A_S| \approx l - l.p_0$ and $|B_S| \approx l.p_0$. As $p_0 \leq 0.5$ it could be shown that $l - p_0 \geq l.p_0$ or in other words, $|A_F| \geq |B_S|$. Therefore the expected number of mutations in A_F, that is around $p_m(l - l.p_0)$, is more than the expected number of mutations in B_S, that is $p_m.l.p_0$. As a result it could be expected that the distance of the mutated off-springs with their parents would be more in the cases where PUC is used. Considering an individual is consisted of two parts A, B the probability that number of mutations in part A is more than part B is,

$$P\big(|mut(A)| > |mut(B)||\big) = \sum_{i=1}^{|A|} \sum_{j=0}^{\min(|B|, i-1)} \binom{|A|}{i} p_m^i (1 - p_m)^{|A|-i} \binom{|B|}{j} p_m^j (1 - p_m)^{|B|-j} \qquad (5)$$

Where p_m is mutation probability, $|mut(A)|$ is the number of mutation in part A, and $|mut(B)|$ is the number of mutation in part B.

It has to be noted that the actual amount of exchange by crossover between two individual is expected to be around $l/4$ on average. It is due to the fact that, $l/2$ of the alleles in two random individuals are expected to be similar. This way, it could be expected that from the remaining $l/2$ loci that their alleles are different $l/2$ could be exchanged, on average, by using uniform crossover. Therefore, it could be inferred that $3/4$ of the parent solutions would not be exchanged. To this end, it could be concluded that the actual amount of exchange in eq.(5) would be more towards getting higher as the size of the similar alleles would be always more than the non-similar ones.

As a result it could be concluded that at least 50% of the times mutation would result in having an offspring that its distance becomes farther to its original parent.

An experiment has been conducted that counted number of times that the hamming distance between the parent and mutated off-springs are more than the distance of the parent and its offspring, resulted from crossover. Figure (4) shows the results of this experiment in comparison to the

expected results based from Eq. (5) over different values for p_0 and p_m.

Figure (5) depicts the results of the same experiment when the parent solutions where generated independent and randomly. Likewise the experiment has been done over different values for p_0 and p_m. The experimental results are compared with the result of the probabilistic model.

Comparing Figure (5) with Figure (4) it could be seen that in Figure (5), representing the real case situation, the probabilities are always more than the ones in Figure (5). As mentioned before, it is due to the fact that any two random solution would have a hamming distance of $l/2$ with each other on average. In other words, $l/2$ of the alleles would be similar with each other. This way using the maximum rate of exchange ($p_0 = 0.5$) actual exchange between two parents would be 25%. Thus 75% of the child solution would be similar to its parent. Obviously, it could be expected that the number of mutations in similar parts would always be more than the remaining 25%.

4. CONCLUSION

In this paper the effect of two crossover operators and inverse mutation operator has been studied with regards to the hamming distance. It has been shown that how the resultant off-springs would be placed on search space by application of crossover and mutation operators. Applying the operators on solutions would result to new offspring solutions that their hamming distance with their original parents is dependent on the exchange rate of the alleles. The maximum distance that any off-spring could get from its parents solutions is l/2 over crossover. However this distance could be more by application of mutation operator. Nevertheless, homogeneity of the parent solutions, in other hand, have direct effect on the distance of the resultant off-springs. In other words, the maximum distance of l/2 could only be achieved if the parents are fully dissimilar. It worth to be noted that, due to the fact that the distance between two off-springs in crossover operator is the same as their parents' distance, crossover operator has no effect on population diversity if a generational approach [7] is used as population replacement strategy. The mutation could either reduce or increase the distance between parents and off-springs depending on the amount of exchange that has been done due to the crossover operator.
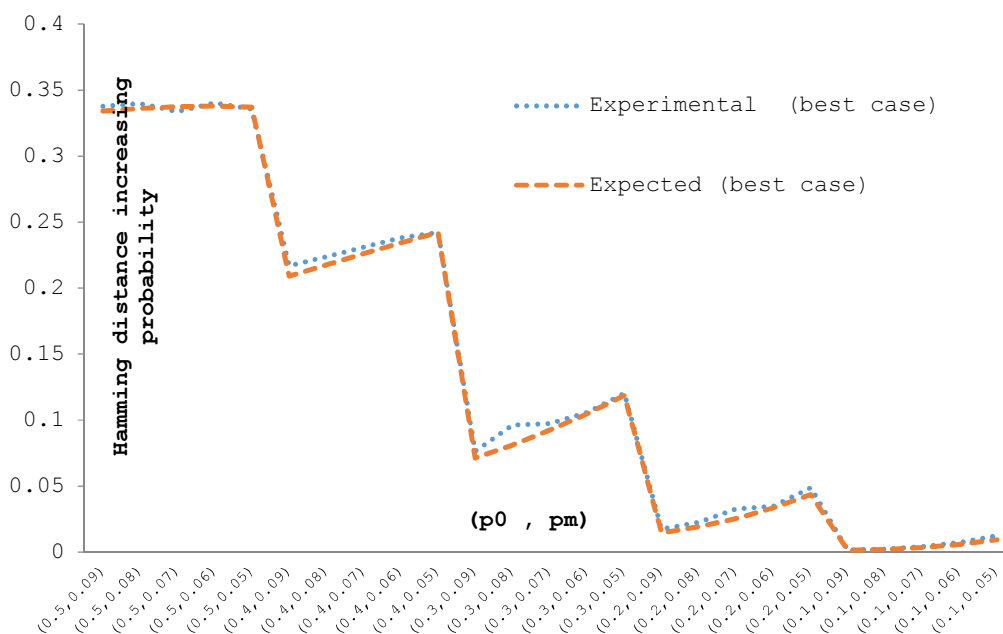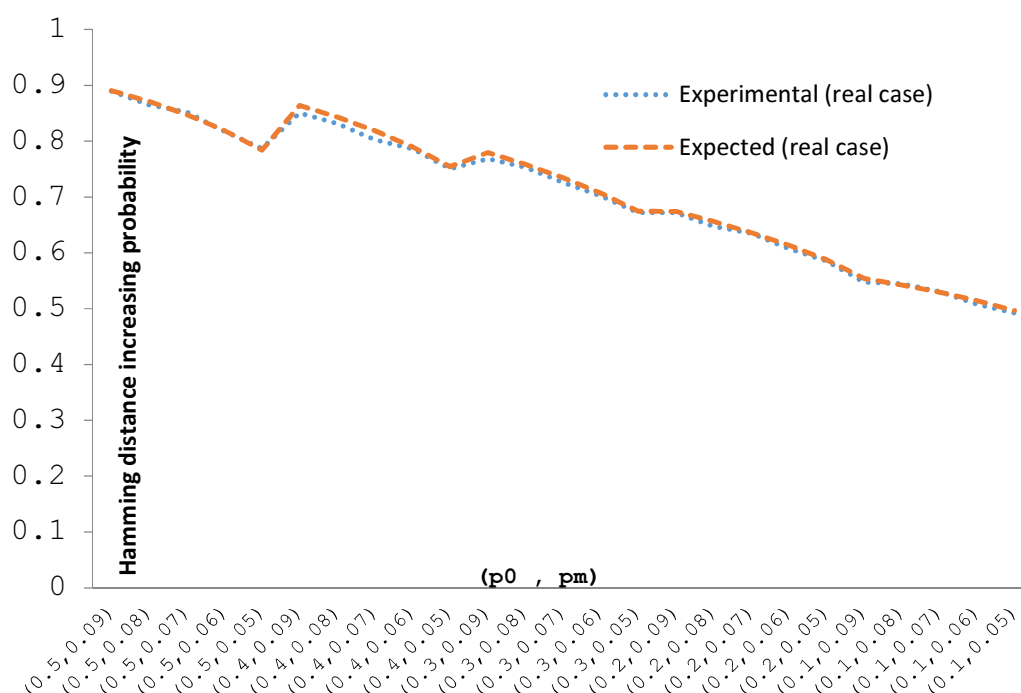

Figure 4: The Probability That The Distance Between Parent And Offspring Increases After Mutation. Experimental Results Are Compared With Mathematical Model Over Different Values For p_0 And p_m. In This Experiment It Was Assumed That Parents Are Negation Of Each Other.

Figure 5: The Probability That The Distance Between Parent And Offspring Increases After Mutation. Experimental Results Are Compared With Mathematical Model Over Different Values Of P0 And Pm. Each Parent Is Generated Randomly.

REFERENCES:

[1] J. H. Holland. Adaptation in Natural and Arti_cial Systems. University of Michigan Press, Ann Arbor, MI, USA, 1975.

[2] Kenneth Alan De Jong. An analysis of the behavior of a class of genetic adaptive systems. PhD thesis, University of Michigan, Ann Arbor, MI, USA, 1975. AAI7609381.

[3] W. Spears and K. De Jong. On the virtues of parameterized uniform crossover. In Rick Belew and Lashon Booker, editors, Proceedings of the Fourth International Conference on Genetic Algorithms, pages 230{236, San Mateo, CA, 1991. Morgan Kaufman.

[4] L.J. Eshelman, R.A. Caruana, and J.D. Scha_er. Biases in the crossover landscape. In Proceedings of the third international conference on Genetic algorithms, pages 10{19. Morgan Kaufmann Publishers Inc., 1989.

[5] Gilbert Syswerda. Uniform crossover in genetic algorithms. In Proceedings of the 3rd International Conference on Genetic Algorithms, pages 2{9, San Francisco, CA, USA, 1989. Morgan Kaufmann Publishers Inc.

[6] Gusz Eiben and Martijn C. Schut. New ways to calibrate evolutionary algorithms. In Patrick Siarry and Zbigniew Michalewicz, editors, Advances in Metaheuristics for Hard Optimization, Natural Computing Series, pages 153{177. Springer, 2008.

[7] Agoston E. Eiben and J. E. Smith. Introduction to evolutionary computing. Natural Computing Series. Springer, 2 edition, December 2010.