CHAPTER 4

PERFORMANCE OF ODD AND EVEN POINT
BASED CROSSOVER

4.1 CONVERGENCE OF ODD AND EVEN POINT BASED
CROSSOVER

The odd and even point crossover is used as the strong exploration
tool in comparison with the other existing crossover operators. It successfully
explored the search space. The algorithm and the results obtained using these
operators are discussed in the previous chapter. The previous chapter proffers
the IR system performance, and the tool properties are not discussed. This
chapter used to analyze the convergence property of the exploration tool namely
odd and even point crossover. That convergence property is vital to caliber the
ability of the optimization tool. Usually convergence of a GA influenced by
three basic operators namely reproduction, crossover and mutation (Marco
Franchini et al 1980). One can’t say that these all three operators have equal
influence over the convergence. Each has it’s own impact. The influence of
these operators can be analyzed by fixing the other two operators. In other
words, the same two operators are not to be used in order to analyze the
influence of the third operator, In our case, interest is shown in exploration that
is crossover only. Hence, the same reproduction is used and mutation
mechanism as that of conventional GA. This arrangement enabled us for the
convergence analysis which is influenced by the crossover. In our case, the
crossover is odd and even point based crossover.
The selection problem for the data fusion in information retrieval has too many combinations. All combinations ought to be explored. This need, leads to the frequent disturbance in the gene structure and results in slow convergence. The price for slow convergence paid over the demand for vigorous exploration as our problem demands it. It will be an apt testing field for our proposed crossover operator. The performance of the proposed operator analyzed in the previous chapter. The important property, that is convergence analysis is going to be carried out in this chapter.

The convergence property analyzed using the following two considerations.

1) Average fitness value of the individual
2) Variation in the average fitness values from one to other generation.

4.1.1 Reason for Selecting These Two Parameter

There are two stopping criteria used in conventional GA. The first one is the fixed number of generations and the second one is the variation in average fitness values among the successive generations. These two has its own merits and demerits. In the first one, the exploration will be stopped after the fixed number of generation even if there is a effective variation among successive generations. This is used as an indicator to indicate the existence of some unexplored areas. If all areas explored, the variation between two successive generations will be minimum.

In the second case, the exact number of generations are not fixed. The variation between the two successive generations used as a indication. If it falls with in the range then the exploration will be stopped. It has its own draw back. During the start, the variation may be minimal. If a weak
exploration tool is found then, the GA will stop it’s exploration during the initial state itself. It means, lots of unexplored areas will be available.

To overcome these two drawbacks, a hybrid termination condition is proposed. In this hybrid termination condition, the number of generation as well as the difference between the average fitness values among the successive generation have been selected.

In any of these three cases, the variation among two successive generations plays a vital role. In the case of fixed number of generation, one has to depend on the knowledge of the expert. But in other two cases it is a different scenario. Hence, the variation in average fitness value between the two successive generations has been selected as the indicator.

4.2 CONVERGENCE ANALYSIS

The convergence analysis has been carried out under two sections. The first one tends to analyze the convergence of conventional GA. The second category tends to analyze the odd and even point crossover based GA. Under these two headings experiments are conducted by varying number of bits used in the encoding. These analyses enabled us to verify the slow convergence of odd and even point crossover based GA.

4.2.1 Convergence Analysis of Conventional GA
Parameters Values

Number of generation : 100
Number of individual per generation : 10
Crossover Probability (P_c) : 0.6
Mutation Probability (P_m) : 0.01

Based on the above set of parameters, experiments are carried out by using two string structure. In the first case, it has been fixed at 12-bits and in second case it raised up to 16 bits. Apart from this there is no variation.

Convergence analysis of 12 bit encoded string under conventional GA

In this convergence analysis number of bits fixed at 12. All other parameter remains the same. The experiments conducted over the same three test collections. The results obtained shown in the following Figures 4.1 to 4.3.

Figure 4.1 Average fitness Values for 100 generations over ADI test collection
The Figures 4.1, 4.2, and 4.3 show the smooth flow in average fitness value. The slow convergence cannot be confirmed unless the difference in average fitness value between two successive generations has been calculated. So, the difference between the successive generations has been calculated and it has been plotted as the step function. The plotted graph which show the difference between the two successive generations are given in Figures 4.4 to 4.6.
Figure 4.4 Difference in precision between two successive generations over ADI test collection

Figure 4.5 Difference in precision between two successive generations over CISI test collection
Convergence analysis of 16 bit encoded string under conventional GA

In this second case the convergence analysis of conventional GA has been carried out with 16 bit encoding. All other parameters remain the same as in the first case. The following Figures 4.7 to 4.9 shows the average fitness value.

Figure 4.6 Difference in precision between two successive generations over CRAN test collection

Figure 4.7 Average fitness values for 100 generation over ADI test collection
The variation between two successive generations has been tabulated and plotted as bar chart. The plotted Figures 4.10 to 4.12 shown below:
Figure 4.10 Difference in precision between two successive generations over ADI test collection

Figure 4.11 Difference in precision between two successive generations over CISI test collection
Figure 4.12 Difference in precision between two successive generations over CRAN test collection

The convergence of conventional GA is always smooth. The same thing has been confirmed once again. The figures 4.1, 4.2, and 4.3, and 4.7, 4.8 and 4.9. show the smooth variation in the average fitness value from one generation to other. In other words, small change in fitness value between successive generations indicates the minimal variation in string structure. The difference between the two successive generations are calculated and plotted in figure no 4.4, 4.5, and 4.6, and 4.10, 4.11, and 4.12. The variation value is also minimum. Hence, it leads the conclusion that, the conventional GA has a slow convergence rate with respect to data fusion problem.

4.2.2 Convergence Analysis of Odd and Even Point Crossover Based GA

In this analysis the convergence of odd and even point crossover GA have been tested. For the compression purpose, the same environment and parameter used in the case of previous subsection are maintained here. The only difference is, the odd and even point crossovers are used. Hence, the analysis has been divided into two categories as in the case of previous chapter.
Convergence analysis of 12-bit encoded string under odd and even point crossover based GA

The average fitness value for each generation over 100 generations has been calculated and tabulated. The graph which plotted based on the tabulated value shown in Figures 4.13 to 4.15.

Figure 4.13 Average fitness values for 100 generation over ADI test collection

Figure 4.14 Average fitness values for 100 generation over CISI test collection
Figure 4.15  Average fitness values for 100 generation over CRAN test collection

The variation between two successive generations calculated were, tabulated and plotted as a bar chart. The plotted Figures 4.16 to 4.18 are shown below.

Figure 4.16 Difference in precision between two successive generations over ADI test collection
Figure 4.17 Difference in precision between two successive generations over CISI test collection

Figure 4.18 Difference in precision between two successive generations over CRAN test collection
The convergence analysis of conventional GA has been carried out with help of the difference between two successive generations. Same procedure has been adapted to analysis the convergence analysis of odd and even point cross over based GA. The figures 4.13, 4.14 and 4.15, show the rapid fluctuation in average fitness value from one generation to other. The figures 4.16, 4.17 and 4.18 show the difference in average fitness value between two successive generations. By comparing the figures 4.16, 4.17, and 4.18 with that of 4.4, 4.5, and 4.6 respectively, the difference in average fitness value of the odd and even point cross over based GA is proved to be higher than that of the conventional GA. Hence, the convergence rate of the conventional GA is faster or smoother compared with odd and even point crossover based GA.

**Convergence analysis of 16 bit encoded string under odd and even point crossover based GA**

In this case, number of bit fixed at 16 and the average fitness value used for convergence analysis. The calculated average fitness value for each generation over 100 generations are calculated and plotted. The plotted Figures 4.19 to 4.21 shown below.
Figure 4.19 Average fitness value for 100 generations over ADI test collection

Figure 4.20 Average fitness value for 100 generations over CISI test collection

Figure 4.21 Average fitness value for 100 generations over CRAN test collection

The difference between two successive generation are plotted in the following Figures 4.22 to 4.24.
Figure 4.22 Difference in precision between two successive generations over ADI test collection

Figure 4.23 Difference in precision between two successive generations over CISI test collection
Figure 4.24 Difference in precision between two successive generations over CRAN test collection

Overall analysis

The previous sub sections showed the variation in fitness value and the difference between the two successive generations. But is won’t serve our purpose. The convergence analyze aims to analysis the convergence property of odd and even point crossover based GA and is attempted to compare it with conventional GA. The results for the above two cases obtained and the comparison part still pending. In order to compare the two cases, a suitable method has to be used. The best known method is the difference between average fitness values. In this case there are two possibilities. The first one is the difference in fitness value among any generations. Second one focus to words the difference between two successive generations.

If the first case is chosen it won’t serve the purpose. This is due to fact that, at the starting and at the end, the difference may completely differ and it can’t be treated as a vigorous exploration. The second case calculated
the difference between two successive generations. Hence it can be treated as an indicator for vigorous exploration. Hence the second case is opted for.

The experiments conducted over 100 times and the results are obtained based on the above mentioned parameter calculated for each trial. The maximum and the minimum difference and the average difference for conventional GA and odd and even point crossover based GA calculated. The following Table 4.1 shows the calculated value for the three test document collection.

**Table 4.1 Maximum and minimum difference between two successive generation**

<table>
<thead>
<tr>
<th></th>
<th>Conventional GA</th>
<th>Odd and Even Point Crossover Based GA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12 bit</td>
<td>16 bit</td>
</tr>
<tr>
<td></td>
<td>MIN</td>
<td>MAX</td>
</tr>
<tr>
<td>ADI</td>
<td>0.000</td>
<td>1.978</td>
</tr>
<tr>
<td>CISI</td>
<td>0.0055</td>
<td>1.737</td>
</tr>
<tr>
<td>CRAN</td>
<td>0.000</td>
<td>0.0842</td>
</tr>
</tbody>
</table>

The table 4.1 shows the maximum and minimum difference between two successive generation for three data collections. The minimum value and the maximum value of the conventional GA is much lower than the odd and even point cross over based GA. Hence, it is concluded that the convergence of conventional GA is smooth when compared with odd and even point crossover based GA.
4.3 CONCLUSION

This chapter analyzes the conventional GA and odd and even point crossover based GA in terms of convergence. The conventional GA has smooth convergence over the odd and even point crossover based GA. The steep variation in odd and even point crossover based GA can be controlled. The method of controlling the convergence has been discussed in the next chapter.