Analysis of Variance for A Influence of Genetic Probability on the Convergence times of Genetic Algorithm

Nie XingXin, Lu CaiWu, Liu ShuXiang and Shen Yang Yang
College of management Xi’an University of Architecture and Technology, China

Abstract: The three-factor about replication, Exchange, mutation probabilities and their interaction is important on the convergence of the genetic algorithm. In order to analysis the effect. This study designed an three-factor cross-group experiment of three probability about reproduction, crossover and mutation, we use analysis of variance on empirical datum to obtain the effect of reproduction, crossover and mutation on convergence of Genetic algorithm. Conclusion of study show that, the influence of reproduction is the most significant, the influence of crossover is the second, the influence of mutation is not significant. But, as the main way of producing new individuality, Reproduction probability is important for getting overall situation optimal solution. In addition, the instrumental error is great which show that the calculate result of Genetic algorithm is randomness.

Key words: Genetic algorithm, variance analysis, Genetic probability, convergence times, crissercross inheritance

INTRODUCTION

Three basic operations of genetic algorithm are reproduction, crossover and mutation which guaranteed searching global optimal solution directionally (Yun et al., 1997). The three Operations dominated by certain probabilities which conclude reproduction probability (P_r), crossover probability (P_c), mutation probability (P_m), whether the three big probabilities choose reasonable or not directly affect the stand or fall of genetic algorithm convergence(Chao, 2007) The long period of study found that different genetic probability have big influence on genetic algorithm, the result mainly displays in: first, converging quickly, but converging to local solution, not the global optimal. Second, converging too slowly, maybe a few hundred generations, even thousands of generations and sometimes simply cannot converge. Therefore, genetic algorithm convergence can not be too short or too long, it must converge to the global optimal solution in the appropriate times (as short as possible).

From the group, in order to measure the good or bad of genetic probability three factors must be considered: convergence times (g), the convergence rate (α), convergence of maximum fitness value (f_max). That is, in the proper convergence times converge to the global optimal solution according to the requirements in the convergence rate. This is also the standard that chooses genetic probability.

The convergence times of genetic algorithm follow lognormal distribution (Su, 2012), but it is effected by many factors, it’s hard to see how much influence genetic probability on it from a variety of measurement results. We can see various factors to the influence degree of the genetic algorithm only by the analysis of variance for genetic probability.

Considering the many factors, the differences of determination value are bound to exist in different factors and different level, this calls variation. The variation is usually caused by the change of various factors and system error. We usually express it by the sum of deviation squares between the determination value x and their average (Yu and Zuo, 2010):

\[ Q = \sum_{i=1}^{n}(x_i - \bar{x})^2 \]  

Meanwhile, the test results are influenced by many factors, so each factors could make corresponding contribution to the total deviation(Yu and Zuo, 2010). Sum of deviation squares is used to represent the errors of test results, then the total sum of deviation squares equal to the sum of sum of deviation squares caused by a single factor, called the additivity of the sum of deviation squares which is the basis of analysis of variance. According to the principle of the additivity of sum of deviation squares, on the basis of sum of deviation squares decomposition and with the aid of F tests, we analyze the influences of the each genetic probability and their interaction on the total sum of deviation squares, this calls variance analysis. The variance analysis is one of the basic method of dealing with many factors experimental data(Chen et al., 2006).

Corresponding Author: Nie XingXin, College of management Xi’an University of Architecture and Technology, China

6807
COMPREHENSIVE EXPERIMENTAL DESIGN BASED ON THREE GENETIC PROBABILITIES CROSS GROUP

Genetic probability includes reproduction probability, crossover probability and mutation probability, so the variance analysis based on comprehensive experimental data of three genetic probabilities cross classification could be used to determine the influences of the three genetic probabilities and their interaction on convergence times of genetic algorithm (Zhang et al., 2005) The experimental arrangement is shown in Table 1.

Each of the sum of deviation squares mainly includes (Yun et al., 1997):

- Total sum of deviation squares $Q_T$
- Sum of deviation squares $Q_s$ caused by the factors $P_r$ change in test level
- Sum of deviation squares $Q_c$ caused by the factors $P_c$ change in test level
- Sum of deviation squares $Q_a$ caused by the factors $P_a$ change in test level
- Sum of deviation squares $Q_{rc}$, $Q_{ca}$, $Q_{ac}$ caused by factors $P_r$, $P_c$, and $P_a$ mutual influence
- Sum of deviation squares $Q_{ak}$, $Q_{ck}$, $Q_{ca}$ mutual influence
- Test error effect $Q_e$

The calculation method of each of sum of deviation squares is as follows:

\[
Q_T = \frac{1}{N} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (T_{ijk} - \bar{X})^2
\]

\[
Q_s = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (X_{ijk} - \bar{X})^2
\]

\[
Q_c = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (C_{ijk} - \bar{C})^2
\]

\[
Q_a = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (A_{ijk} - \bar{A})^2
\]

\[
Q_{rc} = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (R_{ijk} - \bar{R})^2
\]

\[
Q_{ca} = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (C_{ijk} - \bar{C})^2
\]

\[
Q_{ac} = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (A_{ijk} - \bar{A})^2
\]

\[
Q_{ak} = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (A_{ijk} - \bar{A})^2
\]

\[
Q_{ck} = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (C_{ijk} - \bar{C})^2
\]

\[
Q_{ak} = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (A_{ijk} - \bar{A})^2
\]

\[
Q_e = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (E_{ijk} - \bar{E})^2
\]

According to the additivity of the sum of deviation squares (Yun et al., 1997):

\[
Q_T = Q_s + Q_c + Q_a + Q_{rc} + Q_{ca} + Q_{ac} + Q_{ak} + Q_{ck} + Q_{ak} + Q_e
\]

The above $r$, $m$, $n$ represent the level for factor $P_r$, factor $P_c$, factor $P_a$, $n$ is the number of repeat test frequency.

Among them:

\[
T = \frac{1}{N} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} T_{ijk}
\]

\[
T_s = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} T_{ijk}
\]

\[
T_c = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} C_{ijk}
\]

\[
T_a = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} A_{ijk}
\]

\[
Q_{rc} = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (R_{ijk} - \bar{R})^2
\]

\[
Q_{ca} = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (C_{ijk} - \bar{C})^2
\]

\[
Q_{ac} = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (A_{ijk} - \bar{A})^2
\]

\[
Q_{ak} = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (A_{ijk} - \bar{A})^2
\]

\[
Q_{ck} = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (C_{ijk} - \bar{C})^2
\]

\[
Q_{ak} = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (A_{ijk} - \bar{A})^2
\]

\[
Q_e = \frac{1}{mn} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (E_{ijk} - \bar{E})^2
\]

The degrees of freedom of each of the sum of deviation squares is: the degrees of freedom of the factor $P_r$ $P_c$ $P_a$ (Jian et al., 2013) are $f_r = r-1$, $f_s = s-1$, $f_c = m-1$; The degrees of freedom of the interaction between factor $P_r$ and $P_t$ (Fallah-Jamshidi et al., 2010), factor $P_t$ and $P_a$ factor $P_c$ and $P_a$ are: $f_{r-t} = (r-1)(c-1)$, $f_{s-t} = (s-1)(m-1)$, $f_{r-c} = (r-1)(c-1)(m-1)$; the degree of freedom of the interaction between all of them is $f_{r-s-t-c} = (r-1)(c-1)(m-1)$ (Tan et al., 2013); the degrees of freedom of the sum of deviation squares
which has deviated effects is \( f_{\text{cox}} = (r-1)(c-1)(m-1) \); the
degrees of freedom of the total sum of deviation squares
is \( f_r - r\text{cmin}-1 \).

VARIANCE ANALYSIS OF SAMPLE DATA

Steps of variance analysis: The steps of variance analysis
based on the data of three genetic probabilities and same
testing times [Bai et al., 2009]:

- Calculate the sum of deviation square \( \sum Q \) according to
  the sample data
- Calculate the degree of freedom \( f \) for each the sum of deviation square \( \sum Q \)
- Calculate each the estimated variance \( S_i^2 \)
- Calculate each F value according to the estimated variance and the degree of freedom;

Choose the significance level \( \alpha \), then look up the
critical value \( F_{\alpha} \) from F distribution table according to the
corresponding degree of freedom;

Compare F with \( F_{\alpha} \). F<\( F_{\alpha} \) means it has no obvious
effect on the convergence of genetic algorithm, \( F_{\alpha} \geq F \)
means it has a obvious effect on the convergence of genetic algorithm. Meanwhile, \( F_{\alpha} \geq F_{0.05} \), means the effect is
of obviousness, \( F_{\alpha} \geq F_{0.01} \), means it has a significant effect
on the convergence of genetic algorithm [Jin and Shan, 2013].

Variance analysis: Variance analysis is showed in
Table 2.

Determination of the test conditions: The parameters are selected within reason, in order to ensure that the genetic
algorithm can converge in normal times. Select 0.04, 0.05
and 0.06 for Reproduction probability \( p_r \), 0.5, 0.55 and 0.6
for crossover probability \( p_c \); 0.005, 0.006 for mutation
probability \( p_m \), then \( r = 3, c = 3, m = 2, 10 \) for repeated test
times in each level, \( n = 10 \).

Result analysis: Convergent times that in different index
and level are showed in Table 3 base on perform simulation.

Because of the convergence time of the genetic algorithm follow lognormal distribution, the sample data
should be translated according to the Eq. 2 before discuss them with analysis of variance:

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of deviation square</th>
<th>Degree of freedom</th>
<th>Estimated variance</th>
<th>F value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_e )</td>
<td>( Q_e )</td>
<td>( f_e = m-1 )</td>
<td>( S_o^2 = Q_o/F_o )</td>
<td>( S_S^2/S_S^2 )</td>
</tr>
<tr>
<td>( P_{ex} )</td>
<td>( Q_{ex} )</td>
<td>( f_{ex} = (r-1)(c-1) )</td>
<td>( S_{ex}^2 = Q_{ex}/F_{ex} )</td>
<td>( S_{ex}^2/S_{ex}^2 )</td>
</tr>
<tr>
<td>( P_{ex}\times P_{ex} )</td>
<td>( Q_{ex} )</td>
<td>( f_{ex} = (r-1)(c-1)(m-1) )</td>
<td>( S_{ex}^2 = Q_{ex}/F_{ex} )</td>
<td>( S_{ex}^2/S_{ex}^2 )</td>
</tr>
<tr>
<td>( P_{ex}\times P_{ex}\times P_{ex} )</td>
<td>( Q_{ex} )</td>
<td>( f_{ex} = (r-1)(c-1)(m-1) )</td>
<td>( S_{ex}^2 = Q_{ex}/F_{ex} )</td>
<td>( S_{ex}^2/S_{ex}^2 )</td>
</tr>
<tr>
<td>Trial error</td>
<td>( Q_e )</td>
<td>( f_r = r\text{cmin}-1 )</td>
<td>( S_e^2 = Q_e/F_e )</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>( Q_e )</td>
<td>( f_r = r\text{cmin}-1 )</td>
<td>( S_e^2 = Q_e/F_e )</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Convergent time in different genetic probability

<table>
<thead>
<tr>
<th>( P_{r0.04} )</th>
<th>( P_{r0.55} )</th>
<th>( P_{r0.6} )</th>
<th>( P_{c0.05} )</th>
<th>( P_{c0.55} )</th>
<th>( P_{c0.6} )</th>
<th>( P_{m0.005} )</th>
<th>( P_{m0.006} )</th>
<th>( P_{m0.008} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pr1(0.04)</td>
<td>Pr2(0.05)</td>
<td>Pr3(0.06)</td>
<td>Pr4(0.05)</td>
<td>Pr5(0.05)</td>
<td>Pr6(0.06)</td>
<td>Pr7(0.05)</td>
<td>Pr8(0.05)</td>
<td>Pr9(0.05)</td>
</tr>
<tr>
<td>43</td>
<td>147</td>
<td>51</td>
<td>32</td>
<td>50</td>
<td>49</td>
<td>84</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>53</td>
<td>52</td>
<td>63</td>
<td>49</td>
<td>32</td>
<td>29</td>
<td>49</td>
<td>35</td>
<td>59</td>
</tr>
<tr>
<td>40</td>
<td>44</td>
<td>81</td>
<td>28</td>
<td>64</td>
<td>44</td>
<td>52</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>34</td>
<td>36</td>
<td>85</td>
<td>32</td>
<td>40</td>
<td>52</td>
<td>27</td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td>61</td>
<td>110</td>
<td>70</td>
<td>74</td>
<td>49</td>
<td>27</td>
<td>35</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>47</td>
<td>110</td>
<td>121</td>
<td>46</td>
<td>46</td>
<td>45</td>
<td>43</td>
<td>71</td>
<td>65</td>
</tr>
<tr>
<td>58</td>
<td>22</td>
<td>56</td>
<td>63</td>
<td>32</td>
<td>76</td>
<td>39</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>53</td>
<td>23</td>
<td>74</td>
<td>22</td>
<td>64</td>
<td>75</td>
<td>33</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>56</td>
<td>62</td>
<td>40</td>
<td>34</td>
<td>27</td>
<td>48</td>
<td>48</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>84</td>
<td>45</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>33</td>
<td>30</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>59</td>
<td>47</td>
<td>88</td>
<td>75</td>
<td>76</td>
<td>28</td>
<td>28</td>
<td>67</td>
<td>134</td>
</tr>
<tr>
<td>107</td>
<td>126</td>
<td>61</td>
<td>48</td>
<td>72</td>
<td>52</td>
<td>46</td>
<td>110</td>
<td>81</td>
</tr>
<tr>
<td>108</td>
<td>61</td>
<td>85</td>
<td>44</td>
<td>28</td>
<td>64</td>
<td>75</td>
<td>78</td>
<td>105</td>
</tr>
<tr>
<td>75</td>
<td>54</td>
<td>77</td>
<td>107</td>
<td>103</td>
<td>106</td>
<td>68</td>
<td>68</td>
<td>71</td>
</tr>
<tr>
<td>45</td>
<td>71</td>
<td>45</td>
<td>120</td>
<td>49</td>
<td>65</td>
<td>94</td>
<td>46</td>
<td>66</td>
</tr>
<tr>
<td>95</td>
<td>120</td>
<td>44</td>
<td>65</td>
<td>58</td>
<td>28</td>
<td>110</td>
<td>110</td>
<td>59</td>
</tr>
<tr>
<td>161</td>
<td>37</td>
<td>164</td>
<td>70</td>
<td>31</td>
<td>67</td>
<td>87</td>
<td>76</td>
<td>37</td>
</tr>
<tr>
<td>80</td>
<td>113</td>
<td>86</td>
<td>38</td>
<td>67</td>
<td>76</td>
<td>51</td>
<td>121</td>
<td>80</td>
</tr>
<tr>
<td>81</td>
<td>75</td>
<td>98</td>
<td>61</td>
<td>41</td>
<td>59</td>
<td>38</td>
<td>30</td>
<td>69</td>
</tr>
<tr>
<td>155</td>
<td>55</td>
<td>50</td>
<td>91</td>
<td>68</td>
<td>51</td>
<td>37</td>
<td>90</td>
<td>46</td>
</tr>
</tbody>
</table>

6809
Table 4: Computation sheet  

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Sum of squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>183.74</td>
<td>33759.50</td>
</tr>
<tr>
<td>T_i</td>
<td>71.41 53.25 59.08</td>
<td>11425.04</td>
</tr>
<tr>
<td>T_j</td>
<td>62.39 59.12 62.22</td>
<td>11259.95</td>
</tr>
<tr>
<td>T_k</td>
<td>75.80 107.94</td>
<td>17386.29</td>
</tr>
<tr>
<td>T_1</td>
<td>24.41 22.37 24.63 18.56 17.96 16.73 19.42 18.79 20.87</td>
<td>3815.45</td>
</tr>
<tr>
<td>T_2</td>
<td>30.88 40.53 21.80 31.44 23.12 35.97</td>
<td>5888.12</td>
</tr>
<tr>
<td>T_3</td>
<td>24.26 38.13 24.35 34.77 37.18 35.04</td>
<td>5811.25</td>
</tr>
</tbody>
</table>

i = 1, 2, 3; j = 1, 2, 3; k = 1, 2

Table 5: Result of three-factor variance analysis 

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of deviation square</th>
<th>Degree of freedom</th>
<th>Estimated variance</th>
<th>F-value</th>
<th>F(n1, n2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_i</td>
<td>2.845134900</td>
<td>2</td>
<td>1.432256746</td>
<td>9.076885569</td>
<td>3.06</td>
</tr>
<tr>
<td>P_j</td>
<td>0.115345911</td>
<td>2</td>
<td>0.056529256</td>
<td>0.358212951</td>
<td>3.06</td>
</tr>
<tr>
<td>P_m</td>
<td>5.739348533</td>
<td>1</td>
<td>5.739348533</td>
<td>36.37295478</td>
<td>3.84</td>
</tr>
<tr>
<td>P_i * P_m</td>
<td>0.242187722</td>
<td>4</td>
<td>0.060546980</td>
<td>0.383714115</td>
<td>2.37</td>
</tr>
<tr>
<td>P_j * P_m</td>
<td>0.115397278</td>
<td>2</td>
<td>0.05698639</td>
<td>0.361037096</td>
<td>3.06</td>
</tr>
<tr>
<td>P_m * P_k</td>
<td>0.303064855</td>
<td>2</td>
<td>0.151542428</td>
<td>0.960395738</td>
<td>3.06</td>
</tr>
<tr>
<td>P_i * P_j * P_m</td>
<td>0.678376841</td>
<td>4</td>
<td>0.219594210</td>
<td>1.391671935</td>
<td>2.37</td>
</tr>
<tr>
<td>Trial error</td>
<td>25.56224722</td>
<td>162</td>
<td>0.15771650</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>38.81674486</td>
<td>179</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ x'_i = \ln(x_i) \] \hspace{1cm} (2)

Analysis of variance are carried out for the translated data, the process and the result are showed in Table 4 or 5.

May see from the result of Table 5:

- The F value of P_m is 36.373>F(0.1,1,162) that means it has a significant effect on the convergence of genetic algorithm.
- The F value of P_i is 9.0769>F(0.05,2,162) that means it has a significant effect on the convergence of genetic algorithm.
- The F value of P_j is 0.3582<F(0.05,2,162) that means it has no obvious effect on the convergence of genetic algorithm.

According to estimated variance, trial errors take up a great proportion in the genetic algorithm that means the genetic algorithm has much randomness. There are many uncontrolled factors in the genetic algorithm is the main reason that causing the random error. For example, choosing mutations, creating the first generation, choosing crossover points, etc.

Using variance analysis proved that mutation probability have the most significant effect on the convergence time of the genetic algorithm. The second is reproduction probability. And the effort of the crossover probability is not noticeable. Although crossover probability has no obvious effect on the convergence time, but it has an important contribution for searching the global optimal solution comes from crossover probability as the primary way to creat a new generation. It mainly displays in the p_i value is two or three amount levels higher than the p_j or p_m value. The rapid creation of new generation is attributed to the higher crossover probability. Studies show that if p_j is small, then the global optimal solution of genetic algorithm is hard to find.

In addition, errors of the estimated variance take up a great proportion in the whole calculation process. The result fully proves the randomness of the calculation in genetic algorithm. So the preferred method to obtain the conclusions is statistical analysis.

ACKNOWLEDGMENT

This study was supported in part by (Shaanxi funding project sponsor: No.E08001,E08003,2012J004).

REFERENCES


