

CO3091 Computational Intelligence and Software Engineering

Coursework 1 – Optimising Algorithms

Author: Mohamed Mubaarak
User: mm698 // 149025232

Deadline 31st October 2017, 23:59

Question 1 : Performing the runs of the evolutionary algorithm.

The evolutionary algorithm was ran a number of times under different combinations of the parameters that are to be investigated in this experiment in order to see if various combinations resulted in varying levels of fitness of the final solution. These parameters are as follows:

- The rate of crossover labelled as **crossOverRate**
- The size of the population labelled as **Alpha**
- The number of parents per generation is referred to as **mu**
- The number of offspring per generation as **Lambda**

Throughout the duration of the investigation, all other factors and variables will remain constant and unchanged. The experiment is conducted numerous times with a configuration file that consists of different values for the cross over rate, population size, number of parents per generation and offspring in each respective generation. This experiment conducts the effects of two varying crossover rates on two populations labelled as **Alpha₁ = mu₁ = lambda₁** and **Alpha₂ = mu₂ = lambda₂**. Below is an outline for the different combinations of the parameters used in each run (which is also stated in each of the required configuration files for each run):

1. (Alpha₁ = mu₁ = lambda₁) = 5 & CrossOverRate = 0.8
2. (Alpha₁ = mu₁ = lambda₁) = 5 & CrossOverRate = 0.42
3. (Alpha₂ = mu₂ = lambda₂) = 137 & CrossOverRate = 0.8
4. (Alpha₂ = mu₂ = lambda₂) = 137 & CrossOverRate = 0.42

In order to perform the runs, a java class named **ConfigParser.java** was created to generate the 4 different configuration files that incorporate the combinations of parameters mentioned above along with a seed that would be used to randomise the selection. This java class uses a DOM parsing method to modify an example template xml configuration file and with the assistance of loops and if statements, functions to navigate to the respective tag locations (e.g. <property name="alpha">5</property>) and modify the parameters with an alternate value that is to be evaluated in each run of the experiment. In addition to this, a second loop was used in which a variable (namely 'i') would increment from 1 to 30 in 4 loops to simulate the usage of a seed, such that, 30 files would be produced for each combination of parameters stated above, each with a different seed. The name scheme used to name the generated configuration files gives an indication of the parameters used to form the file itself.

For example here is a random configuration file that's generated as a result and it is named "**conf13708_17.txt**". What this means is that this is a configuration file that uses 137 for its **population value (Alpha₂ = mu₂ = lambda₂)**, 0.8 for its **CrossOverRate** and, in this example, a **seed** of 17. This example in particular uses the third combination of parameters used in the experiment. The same naming scheme is used for the 120 configuration text files that are generated with a single execution of the java class ConfigParser.java.

Next, a short bash script named "**script.sh**" functions to take every file generated in the configurations folder/directory and runs each configuration (with debug echo output stating which file is being ran in command-line) into OPT4J which is the tool that is going to be used to run each iteration of the algorithm along with its configuration. The configuration files are generated in such a way that when they are ran in OPT4J are designed to create output files as results of running the algorithm with the combination of parameters and a random seed into a folder named "outputs" as tab separated values (.tsv) following the same naming scheme used to name the configuration files. This generates a total of 120 output files since only one configuration is able to generate output for one set of results consisting of 1000 iterations per file, as outlined in every configuration file with the tag "<property name="generations">1000</property>".

The final script is named "**finalResultsScript.sh**" and it aims to gather all 120 output files and produce 4 final result files in the form of comma separated values (*.csv) which consist of the solutions with the best fitness levels and are used to rank the quality of potential solutions generated by the evolutionary algorithm.

Question 2 : Performing statistical tests

The most suitable statistical test for the project is the **Friedman Test**. The reason for this is that the experiment deals with more than 2 groups where the results/solutions form paired data. Here, paired/related data means that there is a one to one relationship between the number of data points in each data set of each group, which in the case of this experiment, is the combination of parameter values. Furthermore, given the objective of this investigation which is **to evaluate whether different parameter values influence the quality of solutions found by an evolutionary algorithm**, the reason the Friedman Test is best suitable for this is because the evaluation of N groups is stronger at detecting significant differences when they exist compared to evaluating two groups per comparison which falls directly in line with the objective of this experiment.

An additional statistical test was conducted on the matrix of results generated from the evolutionary algorithm is the Post-Hoc test for the Friedman Test known as the **Nemeyi Test**. One advantage of running a Post-Hoc statistical test is that it is able to outline which pair of groups are significantly different. However, a disadvantage for a Post-Hoc test is that they are considered weaker than a regular Friedman or Kruskal-Wallis test for example. This means that they are not as good at detecting the existence of a significant difference between groups even if a significant difference is detected in an N group test conducted on the same groups

Hypotheses

➤ (H0) Null Hypothesis: Group 1 = Group 2

There is no significant difference in the fitness levels of solutions generated by two groups with different rates of crossover and population size.

➤ (H1) Alternate Hypothesis: Group 1 \neq Group 2

There exists at least one pair of groups where there is a significant difference in the fitness levels of solutions.

The null hypothesis is to be accepted upon conducting all the runs in this experiment and discovering no significant difference in the quality of solutions produced by the evolutionary algorithm for main 4 combinations of parameter values. It indicates that modifying the crossover rate and population size has little to no impact in solutions produced by the evolutionary algorithm and vice versa for the alternate hypothesis.

The groups for this experiment are as follows:

- Group 1 (05-08) v Group 2 (05-042)
 - Group 1 (05-08) v Group 3 (137-08)
 - Group 1 (05-08) v Group 4 (137-042)
 - Group 2 (05-042) v Group 3 (137-08)
 - Group 2 (05-042) v Group 4 (137-042)
 - Group 3 (137-08) v Group 4 (137-042)
- (Where $XX = (\text{Alpha} = \mu = \text{lambda})$, and $YY = \text{CrossOverRate}$ in Group ($XX - YY$))

The groups mentioned above were stored into a matrix of data values and the Friedman Test was conducted on this matrix to make the adequate comparisons between pairs of groups. The results of the Friedman Test are as follows:

$\text{friedman chi-squared} = 83.56, df = 3, p\text{-value} < 2.2e-16$

After the Nemeyi test was conducted on the matrix of groups (**posthoc.friedman.nemeyi test(m)**) the following observations were made about the levels of significant difference between the different groups:

	Group1	Group2	Group3
Group2	0.3786	-	-
Group3	1.3e-10	4.4e-14	-
Group4	0.0012	6.9e-07	0.0144

The statistical tests indicate a variety of results in relation to significant difference. Firstly, under the Friedman test returned a P value of 2.2e-16 which is smaller than the level of significance for the experiment which is set to 0.05. This means that based on the Friedman Test the null hypothesis can be rejected as it can be assumed that the data is not considered similar or alike; the result is **statistically significant**.

However, under the Post-Hoc Nemeyi test, the table produced above outlines that a variety of P-values were produced both greater than the significance level and lower than the significance level. For example, when comparing Group 1 and Group 2, the **P-value produced (0.3786) is greater than the significance level indicating that the results are not statistically significant**, and so the null hypothesis is accepted for this instance. It is important to note that this was the only comparison that resulted in an outcome that was not statistically significant in this experiment. Furthermore, the remaining comparisons in the matrix produce a P-value which is **lower than the significance level of 0.05** which indicates that the results are **statistically significant** leading to the null hypothesis being rejected.

Question 3 : Conclusions

The conclusions gathered as a result of the runs are that increasing the Crossover-Rate has no significant difference in the fitness level of solutions generated by the evolutionary algorithm. This is indicative in the comparison between results0508 (Group 1) and results05042 (Group 2) in the matrix. The P-value produced by this comparison is greater than the set significance level of 0.05 thus accepting the null hypothesis. The best fitness results generated in each of the groups' files show a variation of at most 800 between the two groups and this difference was produced by just modifying the crossover-rate and keeping the population size ($\text{Alpha} = \mu = \text{lambda}$) constant between both groups.

Alternatively comparing results05042 (Group 2) and results 137042 (Group 4) where the Crossover-Rate stayed the same whilst the population size was changed resulted in a P-value of 6.9e-07 which is greatly smaller than the level of significance of 0.05. In the best results fitness files produced for each of these configurations, the variation between the data values was **significantly** drastic; in some cases a difference nearing 4000 between paired data sets.

Prior to conducting the experiment, the initial thoughts were that an increased Crossover-Rate would have a significant effect on the fitness level of solutions generated by the evolutionary algorithm since there are more crossovers occurring at any given time per/between generation(s). This proved not to be the case as certain groups where the Crossover-Rate increased did not show a greater difference in significance level. An example of this is Group 3 (13708) and Group 4 (137042) where although the P-value returned (0.0144) was lower than 0.05, the variation in the results' fitness levels between both groups were only in the hundreds as opposed to the thousands; which was the case in instances where the two groups being compared maintained a shared value for the Crossover-Rate, whilst the population size was modified as mentioned above.