Chris Peabody

cgpzbd@mst.edu

COMP SCI 5401 FS2015 Assignment 2c

Methodology:

In this EA, two sets of controllers were co-evolved. The first set was to control Mrs. Pac-man, and the second set was to control the ghosts which were to face her. One ghost controller at a time was used to control all three ghosts in a single play through of the game. These controllers were evolved using genetic programming trees.

Both the Mrs. Pac-man controllers and the ghost controllers had access to the same functional nodes. These were addition, subtraction, multiplication, division, and random (which would return a float value between the two input numbers). They would each have two children, and would perform their given functionality when evaluated.

The two controller types had their own set of terminal nodes, unique to them. Mrs. Pac-man controllers had a node returning the distance from her to the nearest ghost, as well as a node returning the distance from her to the nearest pill. The ghost controllers had a node returning the distance from them to Mrs. Pac-man, and another returning the distance to the nearest other ghost. In addition to these, both controllers had a constant terminal node, which would be given a value between 0 and 2 when initialized.

Experimental Setup:

Several parameters for the experiments were consistent over each test. Those parameters are listed below.

Height of 15
Width of 10
Pill density of 50%
Randomized seed
30 runs of 2000 evaluations each
A 'depth max' of 10. This value is the initial maximum depth a tree, and is also the value at which parse pressure kicks in.
Parent selection for both populations is set to proportional. Survival selection is set to truncation for both.
The parse pressure coefficient is set to 10.

The remaining parameters are what varied for each experiment.

Experiment 1:

Recombination vs mutation chance of 50-50
Mrs. Pac-man lambda of 20 and mu of 10
Ghost lambda of 20 and mu of 10

Experiment 2:

Recombination vs mutation chance of **70-30**
Mrs. Pac-man lambda of 20 and mu of 10
Ghost lambda of 20 and mu of 10

Experiment 3:

Recombination vs mutation chance of 50-50
Mrs. Pac-man lambda of 20 and mu of 10
Ghost lambda of **10** and mu of 10

The first experiment was a sort of basis to see how the other two differ. The second experiment was to see how using mostly recombination would affect the run of the program. The third experiment was to see how cutting the population size of the ghosts in half would affect the run of the program.

Results:

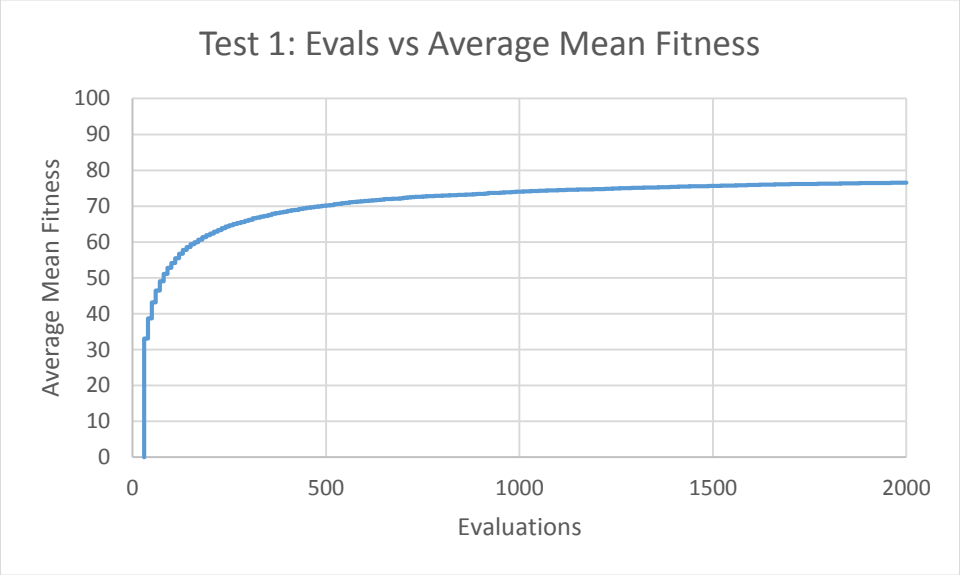

Table 1: In this table the graphical results for the first experiment are shown. The mean fitness values for each generation in each run are averaged together to make a single curve.

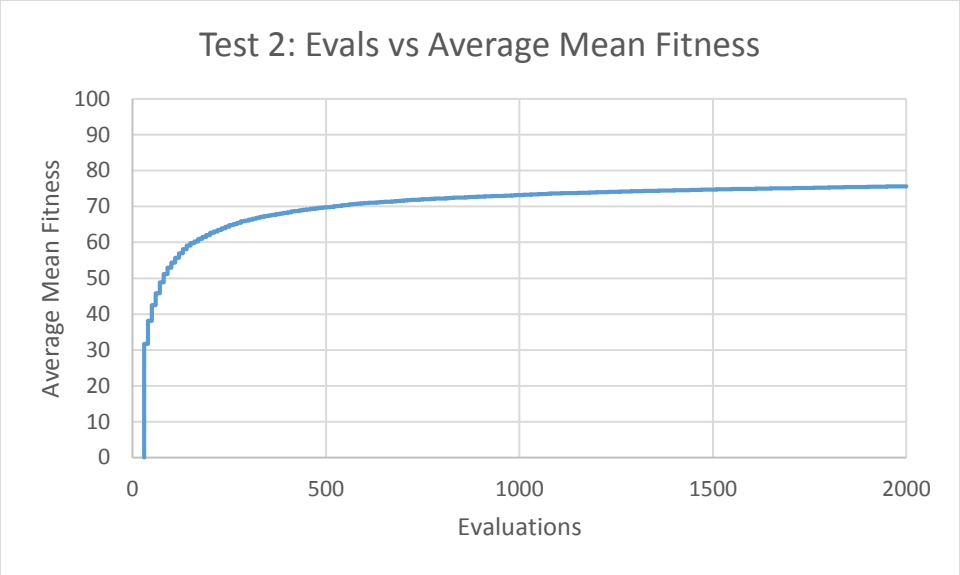

Table 2: In this table the graphical results for the second experiment are shown. The mean fitness values for each generation in each run are averaged together to make a single curve.

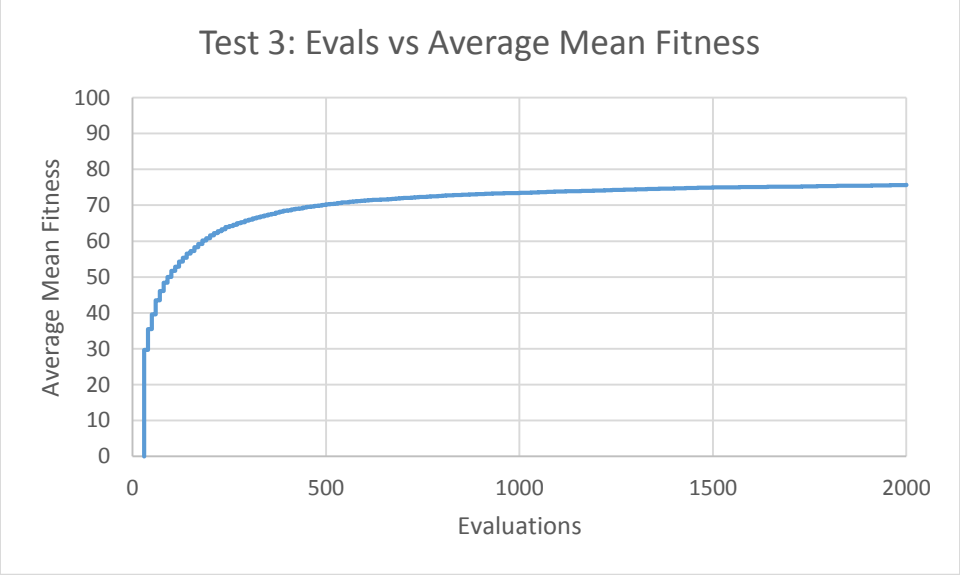

Table 3: In this table the graphical results for the third experiment are shown. The mean fitness values for each generation in each run are averaged together to make a single curve.

Test 1 vs Test 2

F-Test Two-Sample for Variances

	Variable 1	*Variable 2*
Mean	80.86667	79.73333
Variance	4.326437	10.68506
Observations	30	30
df	29	29
F	0.404905	
P(F<=f) one-tail	0.008786	
F Critical one-tail	0.5374	

t-Test: Two-Sample Assuming Equal Variances		
	Variable 1	*Variable 2*
Mean	80.86667	79.73333
Variance	4.326437	10.68506
Observations	30	30
Pooled Variance	7.505747	
Hypothesized Mean Difference	0	
df	58	
t Stat	1.602162	
P(T<=t) one-tail	0.057276	
t Critical one-tail	1.671553	
P(T<=t) two-tail	0.114553	
t Critical two-tail	2.001717	

Table 4: Above is the statistical analysis for the local best results in each run of Test 1 and Test 2. The statistical analysis concluded that the two sets of data have no meaningful difference, and it cannot be said that one test's results were better than another.

Test 1 vs Test 3

F-Test Two-Sample for Variances		
	Variable 1	*Variable 2*
Mean	80.86667	80.13333
Variance	4.326437	6.464368
Observations	30	30
df	29	29
F	0.669275	
P(F<=f) one-tail	0.142666	
F Critical one-tail	0.5374	

t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	*Variable 2*
Mean	80.86667	80.13333
Variance	4.326437	6.464368
Observations	30	30
Hypothesized Mean Difference	0	
df	56	
t Stat	1.222743	
P(T<=t) one-tail	0.113274	
t Critical one-tail	1.672522	
P(T<=t) two-tail	0.226548	
t Critical two-tail	2.003241	

Table 5: Above is the statistical analysis for the local best results in each run of Test 1 and Test 3. The statistical analysis concluded that the two sets of data have no meaningful difference, and it cannot be said that one test's results were better than another.

Test 2 vs Test 3

F-Test Two-Sample for Variances		
	Variable 1	*Variable 2*
Mean	79.73333	80.13333
Variance	10.68506	6.464368
Observations	30	30
df	29	29
F	1.652916	
P(F<=f) one-tail	0.091	
F Critical one-tail	1.860811	

t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	*Variable 2*
Mean	79.73333	80.13333
Variance	10.68506	6.464368
Observations	30	30
Hypothesized Mean Difference	0	
df	55	
t Stat	-0.52905	
P(T<=t) one-tail	0.29945	
t Critical one-tail	1.673034	
P(T<=t) two-tail	0.5989	
t Critical two-tail	2.004045	

Table 6: Above is the statistical analysis for the local best results in each run of Test 2 and Test 3. The statistical analysis concluded that the two sets of data have no meaningful difference, and it cannot be said that one test's results were better than another.

Discussion:

In Tables 1-3, the results from the three experiments can be seen in graphical form. On the x-axis of each graph is shown the evaluation number at any given point. The y-axis shows the average of the mean population fitness in each of the 30 runs of that experiment. The three graphs have very comparable curves, but the statistical analysis is what shows if the data sets actually vary.

Tables 4-6 show the f-tests and t-tests for comparing each of the experimental configurations to one another. For each experiment, the local bests at the end of each of the 30 runs were collected to be used as the experiment's dataset. For each comparison (1 v 2, 1 v 3, 2 v 3) the statistical analysis concluded that there was no statistical meaning in the differences between each data set. In other words, no one set of data was better than another.

Conclusion:

Unfortunately, it appears that the changes that were made in configuration for each of the three experiments were not great enough to have a significant effect on the results produced by the program. The three sets of local-best scores were statistically the same as one another, meaning comparable results were produced by each of the configurations.