

# Analysis of the Effects of Enhanced Selection Concepts for Genetic Programming Based Structure Identification Using Fine-Grained Population Diversity Estimation

Stephan M. Winkler  
Upper Austria University of Applied Sciences  
Heuristic and Evolutionary Algorithms Laboratory  
Softwarepark 11  
A-4232 Hagenberg, Austria  
stephan@heuristiclab.com

Michael Affenzeller  
Upper Austria University of Applied Sciences  
Heuristic and Evolutionary Algorithms Laboratory  
Softwarepark 11  
A-4232 Hagenberg, Austria  
michael@heuristiclab.com

Stefan Wagner  
Upper Austria University of Applied Sciences  
Heuristic and Evolutionary Algorithms Laboratory  
Softwarepark 11  
A-4232 Hagenberg, Austria  
stefan@heuristiclab.com

## ABSTRACT

In this paper we use a formalism for estimating the structural similarity of formulas for measuring the genetic diversity among GP populations. As we show in the results section of this paper, population diversity differs a lot in the test runs depending on the selection schemata used; especially the use of strict offspring selection has a significant effect on the progress of the population's diversity.

## Categories and Subject Descriptors

D.2.8 [Software Engineering]: Metrics—*Process metrics*;  
I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search

## General Terms

Algorithms, Measurement, Design, Experimentation

## Keywords

Genetic Programming, Population Dynamics, Data Mining, Machine Learning, System Identification

## 1. MEASURING GENETIC DIVERSITY IN GP POPULATIONS

When it comes to analyzing GA and GP populations, population diversity is one of the most important aspects since it allows statements about progress and convergence of populations. The similarity of solutions is more difficult to quantify in GP than in traditional GAs due to the fact that traditional measures (for example the Hamming distance) cannot be applied directly because solution candidates do not necessarily have the same length or structure. A very comprehensive overview of program tree similarity and diversity measures has been given for instance in [2].

For our empirical tests analyzing population diversity in GP using different selection concepts we have decided to use fine-grained similarity measurements for GP-based structure identification as described in [4]. The method applied

systematically collects all pairs of ancestor and descendant nodes and information about the properties of these nodes. Additionally, for each pair also the distance (w.r.t. the level in the model tree) and the index of the ancestor's child tree containing the descendant node are collected. The similarity of two models is then calculated by comparing all pairs of ancestors and descendants in one model to all pairs of the other model and averaging the similarity of the respective best matches.

In the context of single-population GP we are mainly interested in the similarity among the individuals of the population. For each model  $m$  of the population  $P$  we calculate the mean similarity with all other individuals in the population:

$$\text{meanSim}(m, P) = \frac{1}{|P| - 1} \sum_{m2 \in P, m2 \neq m} \text{sim}(m, m2) \quad (1)$$

The mean values of all individuals' similarity values are used for calculating the mean similarity measures for populations:

$$\text{meanSim}(P) = \frac{1}{|P|} \sum_{m \in P} \text{meanSim}(m, P) \quad (2)$$

## 2. EMPIRICAL TESTS

Within the last few years a GP based structure identification framework which also uses these further developed selection principles has been implemented for HeuristicLab<sup>1</sup> [3], an open source framework for prototyping and analyzing optimization techniques.

For testing the population diversity analysis method described in the previous section and illustrating graphical representations of the results of these tests we have used the following three data sets: The *NO<sub>x</sub>* data set which contains the measurements taken from a 2 liter 4 cylinder BMW diesel engine at a dynamical test bench (simulated vehicle: BMW 320d Sedan) as well as the *Thyroid* and the *Wisconsin* data sets which are taken from the UCI Machine Learning Repository<sup>2</sup>. Please see for example [1] for details about these data sets.

<sup>1</sup><http://dev.heuristiclab.com>

<sup>2</sup><http://www.ics.uci.edu/~mllearn/>

**Table 1: GP Test Strategies.**

Strategy	Properties
(A) <i>Standard GP</i>	Tournament parents selection ( $k = 3$ )
(B) <i>Standard GP</i>	Proportional parents selection
(C) <i>GP with OS</i>	Offspring Selection [1] Success ratio: 0.8 Comparison factor: 0.8 Maximum selection pressure: 50
(D) <i>GP with strict OS</i>	Strict offspring selection Success ratio: 1.0 Comparison factor: 1.0 Maximum selection pressure: 100

We have tested GP with populations of 1,000 solution candidates (with a maximum tree size of 50 and a maximum tree height of 5), standard subtree exchange crossover, structural as well as parametric node mutation and total 15% mutation rate; the mean squared errors function was used for evaluating solutions, and the maximum number of generations was set to 4000. Other essential parameters vary as summarized in Table 1.

In Table 2 we give the average population diversity values calculated using Equation 2; as each test series was executed several times, we give the average and standard deviation values (written in italic letters).

As we see in the first row, the average similarity values are approximately in the interval [0.2; 0.25] at the beginning of the GP runs, i.e. after the initialization of the GP populations. In standard GP, as can be seen in the first column, the average similarity reaches values above 0.7 after 400 generations and stays at approximately this level until the end of the execution of the GP process. Analyzing the second and the third column we notice that this is not the case in test series B and C: In test series B the similarity values do not rise as high as in series A, and also in test series C we measured significantly lower similarities than in series A.

The highest similarity values are documented for test series D using maximally strict offspring selection: As is summarized in the far right column, during the whole evolutionary process the mutual similarity among the models increases steadily, while also the selection pressure increases. In the end, when the selection pressure reaches a high level (in these cases, the predefined limit was set to 100) and the algorithm stops, we see a very high similarity among the solution candidates, i.e., the population has converged and evolution is likely to have got stuck.

### 3. CONCLUSION

Several variations of genetic programming using different types of selection schemes have been tested using fine-grained similarity estimation. The test results presented in this paper show that population diversity differs a lot in the test runs depending on the selection schemata used. Strict offspring selection in GP enables the production of very high quality models and also gives an intuitive criterion for stopping the algorithm by means of the current selection pressure.

**Table 2: Test results: Population diversity (average similarity values;  $\mu$ ,  $\sigma$ ).**

<i>NO<sub>x</sub> tests</i>					
Gen.	A	B	C	Gen.	D
0	0.247 <i>0.041</i>	0.250 <i>0.031</i>	0.270 <i>0.037</i>	0	0.197 <i>0.039</i>
100	0.723 <i>0.073</i>	0.491 <i>0.051</i>	0.517 <i>0.038</i>	10	0.397 <i>0.039</i>
400	0.813 <i>0.035</i>	0.497 <i>0.058</i>	0.564 <i>0.059</i>	20	0.603 <i>0.049</i>
1000	0.859 <i>0.021</i>	0.510 <i>0.055</i>	0.520 <i>0.052</i>	40	0.810 <i>0.039</i>
4000 (End of run)	0.871 <i>0.019</i>	0.518 <i>0.059</i>	0.526 <i>0.053</i>	End of run	0.985 <i>0.032</i>
<i>Thyroid tests</i>					
Gen.	A	B	C	Gen.	D
0	0.206 <i>0.041</i>	0.205 <i>0.040</i>	0.208 <i>0.036</i>	0	0.197 <i>0.040</i>
100	0.581 <i>0.047</i>	0.241 <i>0.043</i>	0.444 <i>0.035</i>	10	0.397 <i>0.039</i>
400	0.737 <i>0.032</i>	0.321 <i>0.058</i>	0.610 <i>0.026</i>	20	0.602 <i>0.049</i>
1000	0.808 <i>0.029</i>	0.341 <i>0.049</i>	0.692 <i>0.031</i>	40	0.810 <i>0.041</i>
4000 (End of run)	0.812 <i>0.038</i>	0.343 <i>0.056</i>	0.701 <i>0.030</i>	End of run	0.975 <i>0.019</i>
<i>Wisconsin tests</i>					
Gen.	A	B	C	Gen.	D
0	0.205 <i>0.040</i>	0.203 <i>0.039</i>	0.243 <i>0.038</i>	0	0.196 <i>0.042</i>
100	0.698 <i>0.088</i>	0.247 <i>0.146</i>	0.606 <i>0.184</i>	10	0.295 <i>0.096</i>
400	0.704 <i>0.113</i>	0.393 <i>0.168</i>	0.729 <i>0.167</i>	20	0.556 <i>0.062</i>
1000	0.711 <i>0.108</i>	0.353 <i>0.166</i>	0.758 <i>0.158</i>	40	0.777 <i>0.033</i>
4000 (End of run)	0.744 <i>0.114</i>	0.610 <i>0.106</i>	0.765 <i>0.149</i>	End of run	0.972 <i>0.021</i>

### 4. REFERENCES

- [1] M. Affenzeller, S. Winkler, S. Wagner, and A. Beham. *Genetic Algorithms and Genetic Programming - Modern Concepts and Practical Applications*. Chapman & Hall / CRC, 2009.
- [2] E. K. Burke, S. Gustafson, and G. Kendall. Diversity in genetic programming: An analysis of measures and correlation with fitness. *IEEE Transactions on Evolutionary Computation*, 8(1):47–62, 2004.
- [3] S. Wagner. *Heuristic Optimization Software Systems – Modeling of Heuristic Optimization Algorithms in the HeuristicLab Software Environment*. PhD thesis, Johannes Kepler University Linz, 2009.
- [4] S. Winkler, M. Affenzeller, and S. Wagner. Fine-grained population diversity estimation for genetic programming based structure identification. In *Proceedings of the Genetic and Evolutionary Computation Conference GECCO 2008*, pages 1435–1436. Association for Computing Machinery (ACM), 2008.