# A Survey and Analysis of Diversity Measures in Genetic Programming

**Edmund Burke**        **Steven Gustafson**[†]        **Graham Kendall**

School of Computer Science & IT
University of Nottingham
Nottingham, UK NG81BB
{ekb | smg | gxk}@cs.nott.ac.uk
[†]corresponding author

## Abstract

This paper presents a survey and comparison of the significant diversity measures in the genetic programming literature. The overall aim and motivation behind this study is to attempt to gain a deeper understanding of genetic programming dynamics and the conditions under which genetic programming works well. Three benchmark problems (Artificial Ant, Symbolic Regression and Even-5-parity) are used to illustrate different diversity measures and to analyse their correlation with performance. The results show that diversity is not an absolute indicator of performance and that phenotypic measures appear superior to genotypic ones. Finally we conclude that interesting potential exists with tracking ancestral lineages.

## 1  INTRODUCTION

Maintaining population diversity in genetic programming (Banzhaf et al., 1998) is referred to as the key in preventing premature convergence and stagnation in local optima (McPhee and Hopper, 1999)(Ryan, 1994)(Ekárt and Németh, 2000)(McKay, 2000)(Rosca, 1995a). Diversity is the amount of variety in the population defined by what genetic programming individuals 'look' like or how they 'perform'. The number of different fitness values (phenotypes) (Rosca, 1995b), different structural individuals (genotypes) (Langdon, 1996), edit distances between individuals (Ekárt and Németh, 2000), and complex and composite measures (McKay and Abbass, 2001)(Keijzer, 1996)(D'haeseleer, 1994) are used as measures of diversity. At the individual level, diversity measures differences between individuals and is used to select individuals for reproduction or replacement (Eshelman and Schaffer, 1993).

In this study, we examine the previous uses and meanings of diversity, compare these different measures on three benchmark problems and discuss the results. As far as the authors are aware, all the significant diversity measures that occur in the genetic programming literature are reported.

The ultimate goal is to determine a good measurement of population diversity and understand the effects of its influence as the evolutionary search progresses. The overall motivation of this study is that a better understanding of diversity and diversity measures will lead to a better understanding of genetic programming and the advantages and disadvantages of employing it in any given situation.

The following sections examine different measures of diversity, how these measures relate to each other and how they relate to the performance of three genetic programming problems. Section 2 describes measures of population diversity and previous methods of promoting diversity in populations. Section 3 describes the experiments. Section 4 presents and discuss results. Section 5 draws some brief conclusions and Section 6 outlines some ideas for future work.

## 2  DIVERSITY MEASURES

Some measures of diversity are intended to quantify the variety in a population and others are used to measure the difference between individuals. The latter type is used to attempt to control or promote high diversity during a run. The following section surveys both measures that provide a quantification of population diversity and methods used to actively promote and maintain diversity within genetic programming.

## 2.1 POPULATION MEASURES

The most common type of diversity measure is that of structural differences between programs. Koza (1992) used the term *variety* to indicate the number of different programs his populations contained. In this measure, two programs are structurally compared, looking for exact matches. Landgon (1996) felt that genotypic diversity was a sufficient upper bound of population diversity as a decrease in unique individuals must also mean a decrease in unique fitness values. More complex genotype measures count subtrees, size, and type and frequencies of nodes.

Keijzer (1996) measures program variety by the number of unique individuals and subtree variety by counting unique subtrees. Population diversity is a ratio of the number of unique individuals over population size and subtree variety is the ratio of unique subtrees over total subtrees. Tackett (1994) also measures structural diversity using unique subtrees and schemata frequencies. D'haeseleer and Bluming (1994) define the frequency of terminals and functions as "genotypical diversity" and fitness case results as "phenotypical diversity", which are correlated within the population for their study of local populations and demes.

When tree representations of genetic programs are considered as graphs, individuals can be compared for isomorphism (Rosca, 1995a) to obtain a more accurate measure of diversity. Determining graph isomorphism, however, is computationally expensive for an entire population. We could count the number of nodes, terminals, functions and other graph properties in a reasonable time and use this n-tuple to determine whether trees are *possible* isomorphs of each other.

McPhee and Hopper (1999) investigate diversity at the genetic level by tagging each node created in the initial generation with a unique *id*. Root parents, the parents whose tree has a portion of another individual's subtree swapped into it during crossover, are assigned new *memid*s, an auxiliary tag that is initially the same value of the *id*. All the nodes from the root down to the crossover point are assigned new *memid*s to indicate that these nodes have one new child. If there is no mutation in the genetic programming system (as here), then no new *id*s will be created after the initial generation, only *memid*s. McPhee and Hopper found that the number of unique *id*s dramatically falls after initial generations and, by tracking the root parents, after an average of 16 generations, all further individuals have the same common root ancestor.

Phenotypic measures compare the number of unique fitness values in a population. When the genetic programming search is compared to traversing a fitness landscape, this measure provides an intuitive way to think of how the population covers that landscape. Other measures could be created by using fitness values of a population, as done by Rosca (1995a) with entropy and free energy. Entropy here represents the amount of disorder of the population, where an increase in entropy represents an increase in diversity. Bersano-Begey (1997) track how many individuals solve which fitness cases. By monitoring the population, a pressure is added to individuals to promote the discovery of different or less popular solutions.

## 2.2 PROMOTING DIVERSITY

Several measures and methods have been used to promote diversity by measuring the difference between individuals. These methods typically use a non-standard selection, mating, or replacement strategy to bolster diversity. Neighborhoods, islands, niches, crowding and sharing from genetic algorithms are common themes to these methods.

Eschelman and Schaffer (1993) use Hamming distances between individuals to select individuals for recombination and replacement to improve over hill-climbing-type selection strategies for genetic algorithms.

Ryan's (1994) "Pygmie" algorithm builds two lists based on fitness and length to facilitate selection for reproduction. The algorithm maintains more diversity and prevents premature convergence. The advantage of this algorithm is that it does not attempt to "over-control" evolution and uses simple measures to promote diversity.

De Jong et al (2001) use multiobjective optimisation to promote diversity and concentrate on non-dominated individuals according to a 3-tuple of *<fitness,size,diversity>*. Diversity is the average square distance to other members of the population, using a specialised measure of edit distance between nodes. This multiobjective method promotes smaller and more diverse trees.

McKay (2000) applies the traditional fitness sharing concept from Deb and Goldberg (1989) to test its feasibility in genetic programming. Diversity is the number of fitness cases found, and the sharing concept assigns a fitness based on an individual's performance divided by the number of other individuals with the same performance. McKay also studies negative correlation and a *root quartic negative correlation* in (2001) to preserve diversity. Ekárt and Németh (2000) apply fitness sharing with a novel tree distance definition and suggest that it may be an efficient measure of struc-

tural diversity.

By surveying previous work using diversity measures, we designed several experiments to determine relationships between different population measures of diversity and how they correlate to the best fitness of a run.

# 3  EXPERIMENTS

In this study we would like to answer two questions: One, how do different measures of diversity relate to each other, and two, how do those measures correlate to the best fitness of a run. Three common problems are used with common parameter values from previous studies. For all problems, a population size of 500 individuals, a maximum depth of 10 for each individual, a maximum depth of 4 for the tree generation half-n-half algorithm and internal node selection probability of 0.9 for crossover is used. Additionally, each run consists of 51 generations, or until the ideal fitness is found.

The Artificial Ant, Regression and Even-5-Parity problems are used. All three problems are typical to genetic programming and can be found in many studies, including (Koza, 1992). The ant problem is concerned with finding the best strategy for picking up pellets along a trail in a grid. The fitness for this problem is measured as the number of pellets missed. The regression problem attempts to fit a curve for the function $x^4 + x^3 + x^2 + x$. Fitness here is determined by summing the squared difference for each point along the objective function and the function produced by the individual. The parity problem takes an input of a random string of 0's and 1's and outputs whether there are an even number of 1's. The even-5-parity fitness is the number of wrong guesses for the $2^5$ combinations of 5-bit length strings.

To produce a variety of run performances, where we consider the best fitness in the last generation, we designed three different experiments, carried out 50 times, for each problem. The first experiment *random* performs 50 independent runs. The experiment *stepped-recombination* does 50 runs with the same random number seed, where each run uses an increasing probability for reproduction and decreasing probability for crossover. Initially, probability for crossover is 1.0, and this is decreased by 0.02 each time, skipping the value of reproduction set to .98 to allow for exactly 50 runs and ending with reproduction probability of 1.0 and crossover probability 0.0. The last experiment *stepped-tournament* is similar but we begin with a tournament size of 1 and increment this by 1

for each run, until we reach a tournament size of 50. In the *random* and *stepped-tournament* experiments, crossover probability is set to 1.0 and the tournament size in *random* and *stepped-recombination* is 7. The *Evolutionary Computation in Java* (ECJ), version 7.0, (Luke, 2002) is used, where each problem is available in the distribution.

In analysing the results, we compare the 50 runs for fluctuations of diversity levels in the different measures and examine the standard deviation across experiments for each problem. Additionally, the Spearman correlation coefficient (Siegel, 1956) is computed, comparing the ranking of a run's performance and diversity measure for that run (also taken from the last generation's population).

The following measures of diversity were introduced previously and are described next as they are collected for each generation in every run.

**Unique Node** *id*: Tag each node with *id:memid* as in (McPhee and Hopper, 1999) and count number of distinct *id*s in each generation.

**Size of Ancestral Pool:** Since each individual has one root ancestor, in any generation each individuals' line of root ancestors can be traced to the initial generation. It is possible to consider the size of the set that is formed by a set of root parents from the initial generation, and then replacing this set with its intersection with the next generation's root parents. A common ancestor exists when the size becomes 1.

**Entropy:** Calculate the entropy of the population as in (Rosca, 1995a). Entropy is represented as, where "$p_k$ is the proportion of the population $P$ occupied by population partition $k$":

$$-\sum_k p_k \cdot \log p_k$$

Here a partition is assumed to be each possible different fitness value, but could be defined to include a subset of values.

**Pseudo-Isomorphs:** Calculate pseudo isomorphs by defining a 3-tuple of <terminals,nonterminals,depth>, for each individual and count the number of unique 3-tuples in each population. Two identical 3-tuples represent trees which could be isomorphic.

**Genotypes and Phenotypes:** Count the number of unique trees for the genotype measure (Langdon, 1996). The number of unique fitness values in a population represents the phenotype measure (Rosca, 1995b).

The Spearman correlation coefficient is computed as

follows (Siegel, 1956):

$$1 - \frac{6 \sum_{i=1}^{N} d_i^2}{N^3 - N}$$

Where $N$ is the number of items (50 runs), and $d_i$ is the distance between each run's rank of performance and rank of diversity in the last generation. A value of -1.0 represents negative correlation, 0.0 is no correlation and 1.0 is positive correlation. For our measures, if we see ideal low fitness values, which will be ranked in ascending order (1=best,...,50=worst) and high diversity, ranked where (1=lowest diversity and 50=highest diversity), then the correlation coefficient should be strongly negative. Alternatively, a positive correlation indicates that either bad fitness accompanies high diversity or good fitness accompanies low diversity.

## 4    RESULTS AND DISCUSSION

Graphs of 50 runs for each of the three experiments and each problem were examined. Graphs for the ant and regression problems are shown in Figures 1-4. The min, max and standard deviation of each measure (including best fitness) were calculated for each run and the Spearman correlation coefficient was calculated for each of the six diversity measures versus run performance, found in Table 1. This study involved 450 runs of 51 generations each, with each population consisting of 500 individuals, or 13,500,000 individual evaluations.

We found relatively stable standard deviations of best fitness in the ant problem experiments (11.8575, 12.9049, 12.0785) but there were large difference in standard deviations of genotype diversity (14.4554, 124.7823, 37.3990). This variation in best fitness is not indicated by the number of unique trees (genotypes): There is a minimum value of 428 and a maximum of 489. This consistently high genotype diversity does not suggest a strong relationship with the varying performance.

Unique node *ids* and root ancestors converge early in each run. This confirms the results found in (McPhee and Hopper, 1999) that genetic-level diversity is lost very quickly, even with widely varied performance, recombination and tournament values. A further study to consider when these measures converge could be an interesting indicator of other diversity or run performance values. In nearly all of the graphs of diversity measures and best fitness, the most dramatic activity occurs when the number of unique *ids* and root ancestors converges. This activity can been seen in
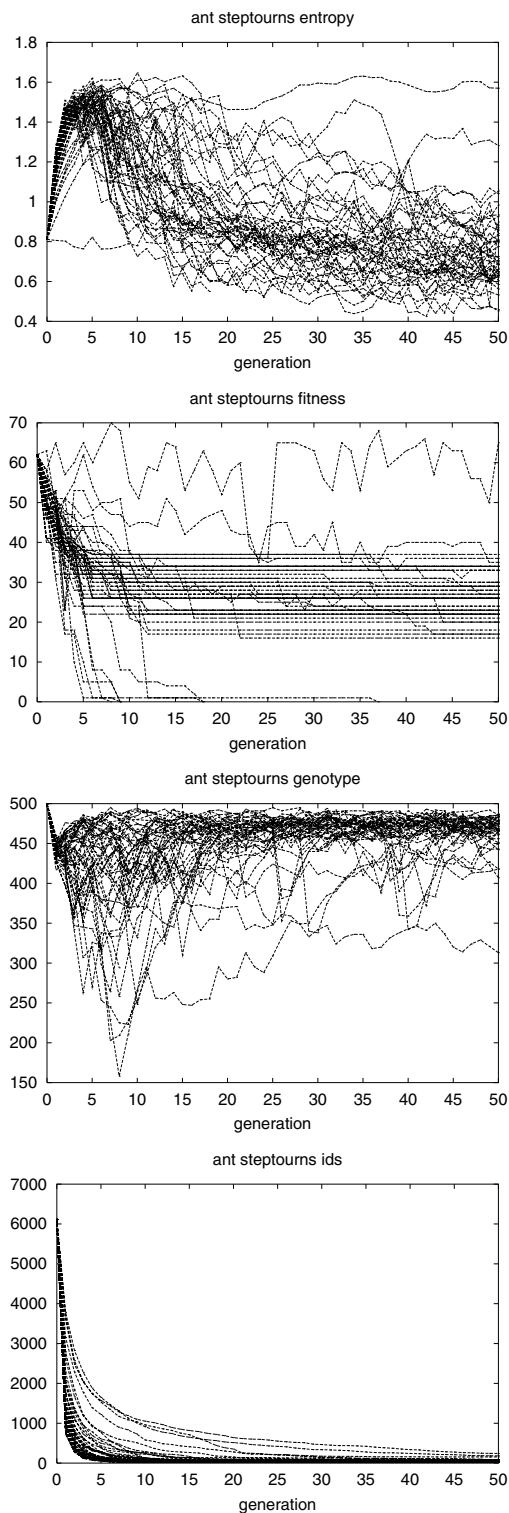


Figure 1: 50 runs of best fitness per generation (top graph) for the ant stepped-tournament experiment. Here, low fitness is better. Also a graph for each of the diversity measures of entropy, genotype, unique node *ids*.
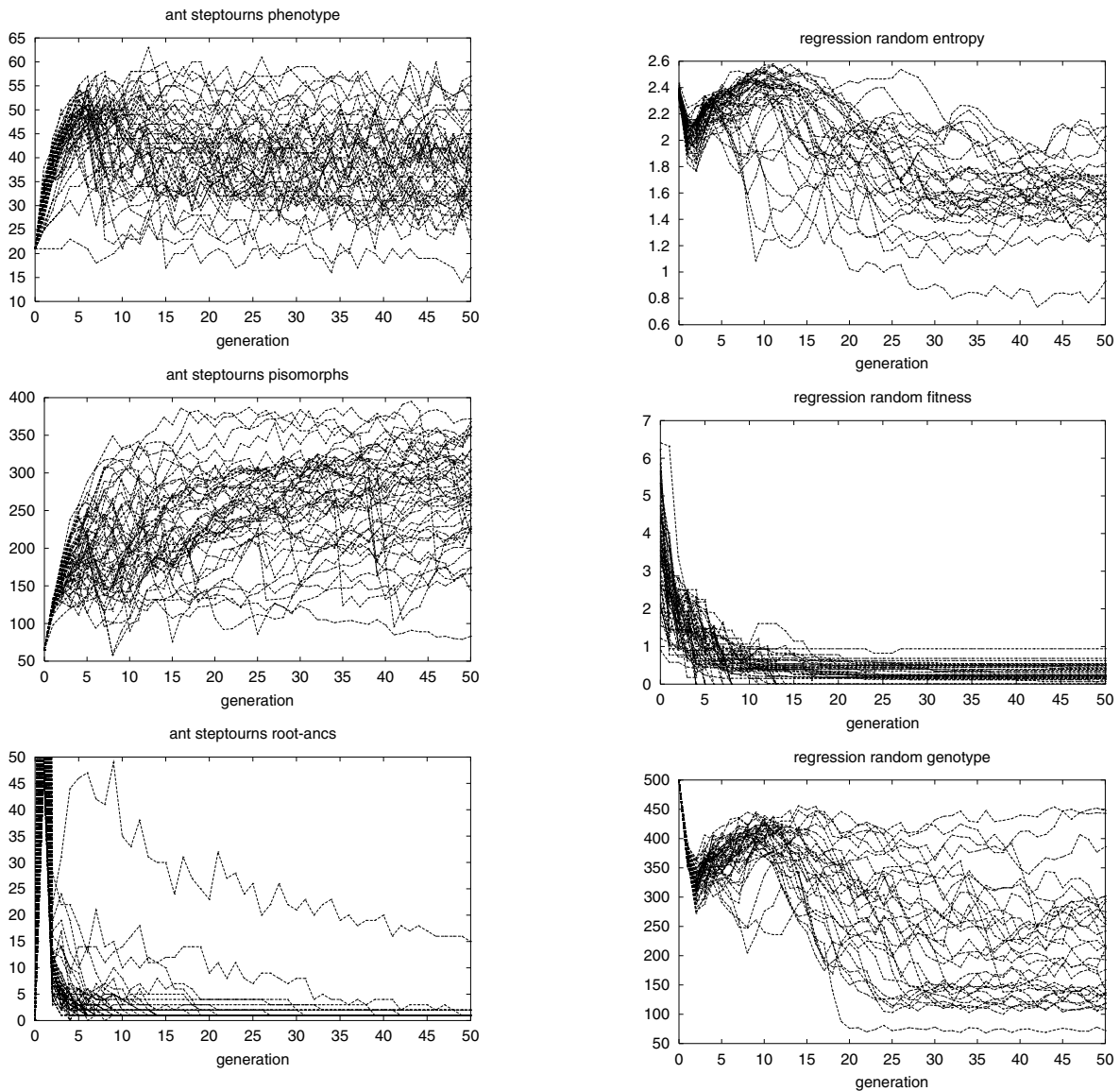
Figure 2: 50 runs of the ant stepped-tournament experiments, showing a graph for each of the diversity measures of phenotype, pseudo-isomorphs, and root ancestors.

Figures 1 through 4. It is not clear, however, how this phenomenon effects evolution and loss of diversity (according to other measures) since, when the number of unique *ids* is reduced and even when a common root ancestor is found, runs are still capable of finding good solutions.

Using the Spearman correlation coefficient we investigated whether runs that produced good fitness had low/high diversity, where ties in ranks were resolved by splitting the rank among tying items (add possible ranks and average). Remembering that negative
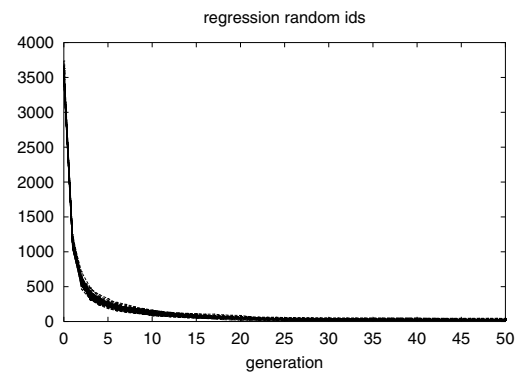
Figure 3: 50 runs of best fitness per generation (top graph) for the regression random experiment. Here, low fitness is better. Also a graph for each of the diversity measures of entropy, genotype, unique node *ids*.

regression random phenotype



regression random pisomorphs
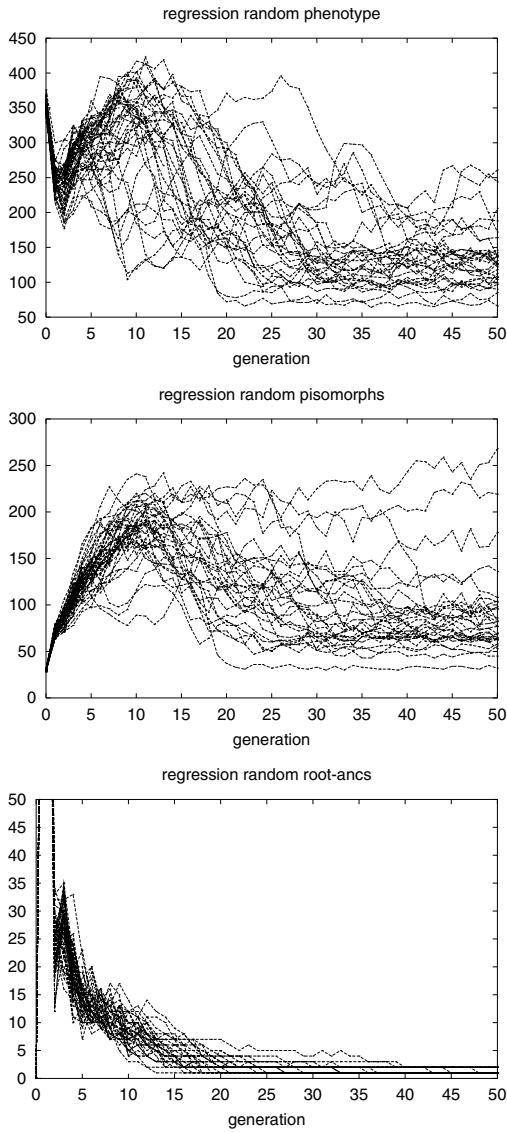


regression random root-ancs



Figure 4: 50 runs of the ant stepped-tournament experiments, showing a graph for each of the diversity measures of phenotype, pseudo-isomorphs, and root ancestors.

correlation (values close to -1.0) suggest that high diversity is correlated with good performance. Table 1 provides the data for all experiments. High negative correlation is seen most consistently with entropy and phenotype diversity. Genotype diversity showed high negative correlation on the regression problem but otherwise varied between little to positive correlation on other problems. While phenotype and entropy always had a negative correlation with performance, values ranged from -0.1608 to -0.8893 with an average correlation of -0.6019 for phenotype and -0.6054 for entropy diversity across all experiments. These were the only

Table 1: Problems ant (a), regression (r) and parity (p) with experiments *random* (rand), *stepped-tournament* (step-t) and *stepped-recombination* (step-r). Values are from the final population. Best fitness ("b.fit") is the best fitness in the final generation. The Spearman correlation coefficient shows perfect correlation with value 1.0 and perfect negative correlation with value -1.0.

| prob. | expr. | col. | spearman | min max | stand.dev |
|---|---|---|---|---|---|
| a | rand | b.fit | | 0.0 39.0 | 11.8575 |
| a | rand | ids | 0.1727 | 25.0 145.0 | 22.4092 |
| a | rand | roots | 0.5014 | 1.0 1.0 | 0.0 |
| a | rand | phene | -0.1608 | 16.0 59.0 | 8.0181 |
| a | rand | gene | 0.4081 | 428.0 489.0 | 14.5543 |
| a | rand | isom | 0.5391 | 121.0 350.0 | 63.3594 |
| a | rand | entro | -0.4195 | 0.4215 1.1566 | 0.1702 |
| a | step-r | b.fit | | 0.0 62.0 | 12.9049 |
| a | step-r | ids | 0.0155 | 15.0 110.0 | 24.1658 |
| a | step-r | roots | 0.1740 | 1.0 4.0 | 0.5291 |
| a | step-r | phene | -0.4088 | 1.0 47.0 | 9.6260 |
| a | step-r | gene | 0.0799 | 1.0 477.0 | 124.7823 |
| a | step-r | isom | 0.3532 | 1.0 348.0 | 83.0020 |
| a | step-r | entro | -0.5590 | -0.0 1.1457 | 0.2160 |
| a | step-t | b.fit | | 0.0 65.0 | 12.0785 |
| a | step-t | ids | 0.2351 | 14.0 242.0 | 42.4240 |
| a | step-t | roots | 0.4253 | 1.0 15.0 | 1.9673 |
| a | step-t | phene | -0.2854 | 17.0 57.0 | 8.9314 |
| a | step-t | gene | 0.3040 | 294.0 488.0 | 37.3990 |
| a | step-t | isom | 0.3394 | 83.0 372.0 | 67.0498 |
| a | step-t | entro | -0.3461 | 0.4525 1.5702 | 0.2155 |
| r | rand | b.fit | | 0.0 0.9399 | 0.2310 |
| r | rand | ids | -0.6552 | 16.0 342.0 | 89.9100 |
| r | rand | roots | -0.6393 | 1.0 21.0 | 5.3113 |
| r | rand | phene | -0.7159 | 66.0 377.0 | 95.6887 |
| r | rand | gene | -0.5779 | 72.0 448.0 | 114.2444 |
| r | rand | isom | -0.5321 | 32.0 268.0 | 53.2196 |
| r | rand | entro | -0.6882 | 0.9297 2.5029 | 0.4044 |
| r | step-r | b.fit | | 0.0 2.8999 | 0.4552 |
| r | step-r | ids | -0.5228 | 4.0 99.0 | 14.9947 |
| r | step-r | roots | 0.0244 | 1.0 8.0 | 1.5133 |
| r | step-r | phene | -0.8703 | 1.0 303.0 | 61.0422 |
| r | step-r | gene | -0.8318 | 1.0 347.0 | 76.7983 |
| r | step-r | isom | -0.8082 | 1.0 165.0 | 36.1054 |
| r | step-r | entro | -0.8430 | -0.0 2.2878 | 0.4713 |
| r | step-t | b.fit | | 0.0 2.8999 | 0.4338 |
| r | step-t | ids | -0.5199 | 8.0 208.0 | 39.7216 |
| r | step-t | roots | -0.0021 | 1.0 16.0 | 3.3859 |
| r | step-t | phene | -0.5797 | 22.0 428.0 | 88.6046 |
| r | step-t | gene | -0.5043 | 28.0 458.0 | 108.1168 |
| r | step-t | isom | -0.4479 | 17.0 249.0 | 49.4191 |
| r | step-t | entro | -0.4001 | 1.0748 2.5894 | 0.3214 |
| p | rand | b.fit | | 3.0 12.0 | 1.9267 |
| p | rand | ids | -0.0142 | 29.0 93.0 | 12.6820 |
| p | rand | roots | 0.5189 | 1.0 1.0 | 0.0 |
| p | rand | phene | -0.6950 | 7.0 16.0 | 1.9489 |
| p | rand | gene | 0.2001 | 422.0 478.0 | 14.2580 |
| p | rand | isom | 0.2635 | 46.0 92.0 | 11.5526 |
| p | rand | entro | -0.6777 | 0.5138 0.9241 | 0.08773 |
| p | step-r | b.fit | | 5.0 14.0 | 2.1462 |
| p | step-r | ids | -0.4573 | 15.0 57.0 | 12.4997 |
| p | step-r | roots | 0.5119 | 1.0 1.0 | 0.0 |
| p | step-r | phene | -0.8119 | 1.0 13.0 | 2.4278 |
| p | step-r | gene | -0.5957 | 1.0 471.0 | 103.8743 |
| p | step-r | isom | -0.0526 | 1.0 111.0 | 18.3605 |
| p | step-r | entro | -0.7039 | -0.0 0.8291 | 0.1801 |
| p | step-t | b.fit | | 1.0 15.0 | 2.6510 |
| p | step-t | ids | 0.2629 | 20.0 225.0 | 32.6593 |
| p | step-t | roots | 0.5934 | 1.0 16.0 | 2.1344 |
| p | step-t | phene | -0.8893 | 3.0 17.0 | 2.5258 |
| p | step-t | gene | 0.4247 | 344.0 485.0 | 28.9229 |
| p | step-t | isom | 0.2311 | 39.0 102.0 | 13.9385 |
| p | step-t | entro | -0.8115 | 0.0445 0.9432 | 0.1325 |

measures that did not show some positive correlation.

Correlation values were not consistently high (statistical significant) but indicate that a relationship may be present. Scatter plots show trends indicated by the Spearman correlation, and Figure 5 shows plots for

the regression problem and *stepped-recombination* experiment. Notice the obvious correlation between low fitness rankings and high diversity rankings for each of the 50 runs for the phenotype, genotype, pseudo-isomorphs and entropy measures. Results suggest that one measure is not definitive but different measures may provide useful information for different problems.

The appearance of consistent negative correlations suggests that better performing runs do have higher diversity. Also confirmed by the correlation study is that the entropy and phenotype measures, and the genotype and pseudo-isomorph measures each have similar results. Since phenotype and pseudo-isomorphs would seem to be less computationally expensive, these measures may be more desirable to track in evolutionary computation systems.

## 5 CONCLUSIONS

The measures of diversity surveyed and compared here demonstrate that the typical genotype measure may not be sufficient to accurately capture the dynamics of a population, which is also suggested in (Ryan, 1994)(Keijzer, 1996).

High variance in performance was not indicated by genotype diversity. The phenotype and entropy measures appear to correlate better with run performance and are less expensive to compute.

The pseudo-isomorph measure appeared to be as informative as genotype diversity and suggests that this simpler measure may be more desirable. Additionally, the consistent early convergence of unique node *ids* and root ancestors, coupled with significant activity in the other measures and performance, show interesting potential for more study.

The relationship between diversity and run performance is not straightforward, and our results indicated some measures had a stronger correlation than others, but not in all experiments and in all problems. This study illustrates the need to carefully define diversity and consider the effects of problem and fitness representation.

## 6 FUTURE WORK

Several extensions to this research were identified and are currently underway. Further experiments on more problems (including real-world) will provide a more thorough investigation. By tracking the convergence of unique *ids*, root ancestors and other measures during evolution, it is hoped that an early indicator for run
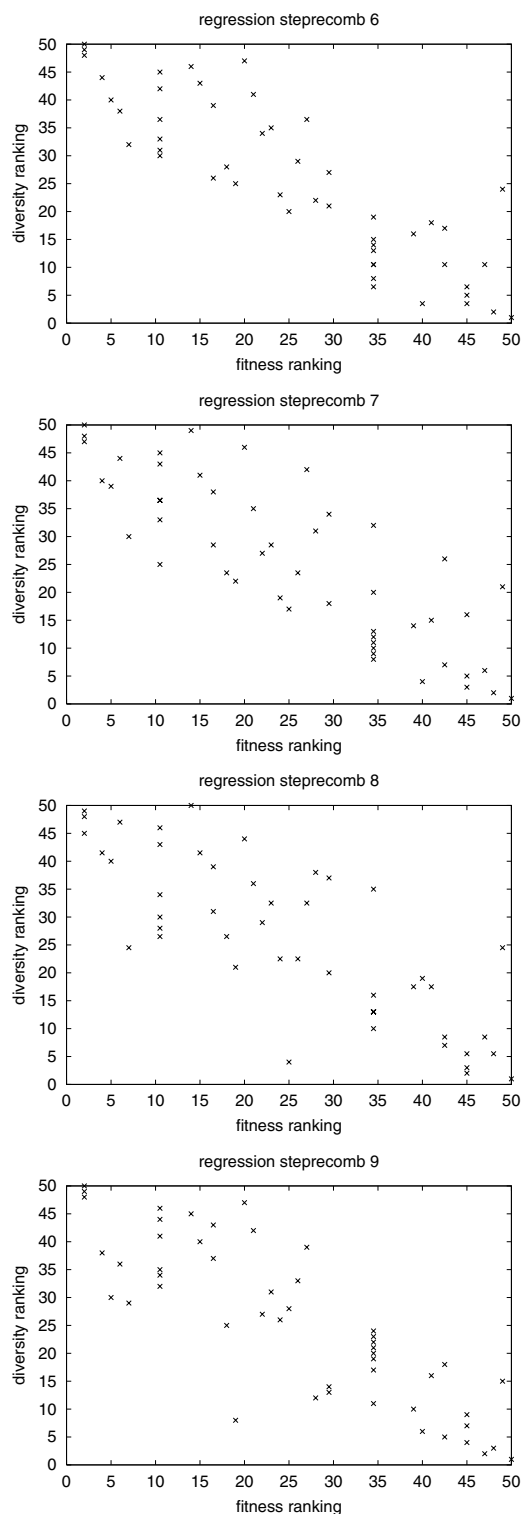


Figure 5: Scatter plots of diversity measures (6=phenotype, 7=genotype, 8=pseudo-isomorphs, 9=entropy) versus best fitness from last generation.

success or failure can be found. Also of interest is using methods to promote diversity and then applying these different diversity measures to determine their effects of improving diversity. Several different and novel diversity measures are also being investigated. The last item of current work examines the computation needed for maintaining the most efficient knowledge (of the evolutionary computation system) to determine effective diversity measures. The research reported is being extended and early experiments indicate that diversity measures based on edit distances provide complimentary and interesting results.

## Acknowledgments

## References

Banzhaf, W., Nordin, P., Keller, R. E., and Francone, F. D. 1998. *Genetic Programming: An Introduction*. Morgan Kaufmann, Inc., San Francisco, USA.

Bersano-Begey, T. 1997. Controlling exploration, diversity and escaping local optima in GP: Adapting weights of training sets to model resource consumption. In J. Koza (ed.), *Late Breaking Papers at the 1997 Genetic Programming Conference*, Stanford University, CA.

de Jong, E., Watson, R., and Pollack, J. 2001. Reducing bloat and promoting diversity using multi-objective methods. In L. Spector et al. (eds.), *Proceedings of the Genetic and Evolutionary Computation Conference*, Morgan Kaufmann, San Francisco, CA.

Deb, K. and Goldberg, D. 1989. An investigation of niche and species formation in genetic function optimization. In J. D. Schaffer (ed.), *Proceedings of the Third International Conference on Genetic Algorithms*, pp 42–50, Washington DC, USA.

D'haeseleer, P. 1994. Context preserving crossover in genetic programming. In *Proceedings of the 1994 IEEE World Congress on Computational Intelligence*, Vol. 1, pp 256–261, IEEE Press, Orlando, FL, USA.

Ekárt, A. and Németh, S. Z. 2000. A metric for genetic programs and fitness sharing. In R. Poli et al. (eds.), *Proceedings of the European Conference on Genetic Programming*, Vol. 1802 of *LNCS*, pp 259–270, Springer-Verlag, Edinburgh.

Eshelman, L. and Schaffer, J. 1993. Crossover's niche. In S. Forrest (ed.), *Proceedings of the Fifth International Conference on Genetic Algorithms*, pp 9–14,

Morgan Kaufman, San Mateo, CA.

Keijzer, M. 1996. Efficiently representing populations in genetic programming. In P. Angeline and K. Kinnear, Jr. (eds.), *Advances in Genetic Programming 2*, Chapt. 13, pp 259–278, MIT Press, Cambridge, MA, USA.

Koza, J. 1992. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, USA.

Langdon, W. 1996. *Evolution of Genetic Programming Populations*. Research Note RN/96/125, University College London, Gower Street, London WC1E 6BT, UK.

Luke, S. 2002. *ECJ: A Java-based evolutionary comp-utation and genetic programming system.* http://www.cs.umd.edu/projects/plus/ecj/.

McKay, R. 2000. Fitness sharing in genetic programming. In D. Whitley et al. (eds.), *Proceedings of the Genetic and Evolutionary Computation Conference*, pp 435–442, Morgan Kaufmann, Las Vegas, Nevada, USA.

McKay, R. and Abbass, H. 2001. Anticorrelation measures in genetic programming. In *Australasia-Japan Workshop on Intelligent and Evolutionary Systems*.

McPhee, N. and Hopper, N. 1999. Analysis of genetic diversity through population history. In W. Banzhaf et al. (eds.), *Proceedings of the Genetic and Evolutionary Computation Conference*, Vol. 2, pp 1112–1120, Morgan Kaufmann, Florida, USA.

Rosca, J. 1995a. Entropy-driven adaptive representation. In J. Rosca (ed.), *Proceedings of the Workshop on Genetic Programming: From Theory to Real-World Applications*, pp 23–32, Tahoe City, California, USA.

Rosca, J. 1995b. Genetic programming exploratory power and the discovery of functions. In J. McDonnell et al. (eds.), *Proceedings of the Fourth Annual Conference on Evolutionary Programming*, pp 719–736, MIT Press, San Diego, CA, USA.

Ryan, C. 1994. Pygmies and civil servants. In K. Kinnear, Jr. (ed.), *Advances in Genetic Programming*, Chapt. 11, pp 243–263, MIT Press.

Siegel, S. 1956. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill Book Company, Inc.

Tackett, W. 1994. *Ph.D. thesis*, University of Southern California, Department of Electrical Engineering Systems, USA.