

Distribution of Degeneracy in the Phenotypic Search Space of Grammar Evolution Algorithms

David Yerger (Moravian College)

Faculty Advisors: Dr. Charles Tolle and Dr. Thomas Montoya
Department of Electrical and Computer Engineering

Abstract

Grammatical evolution has been used to solve optimization problems and to mathematically model dynamic systems. A key element of grammatical evolution is its grammar. However, a method to quantitatively analyze the degeneracy of grammars has not been proposed. The purpose of this study is to explore the degeneracy of common grammars and to show the distribution of phenotypes in the phenotypic search space. MATLAB was used to program grammars, generate genes, collect similar phenotypes, and display the distribution of phenotypes. This research has produced a new method for analyzing degeneracy and shown the uneven distribution of phenotypes.

Introduction

Grammatical evolution uses grammars to translate strings of integers, called genes, into mathematical expressions, called phenotypes. Grammars are normally given in Backus-Naur form (BNF), which defines the method of translating genes into phenotypes. Figure 1 shows an example of a grammar in BNF.

Degeneracy occurs when 2 or more genes are translated into the same phenotype. In this research, the assumption that the phenotypic search space is evenly distributed has been tested.

Procedure

In order to study degeneracy, grammars were programmed using MATLAB. The grammars were then used to translate genes into phenotypes. Further programming in MATLAB compared the phenotypes to form groups of degenerate phenotypes. Each group was given a degeneracy value, and a histogram was produced for each run.

The variables considered were grammar used, gene length, wrapping, number of non-terminals, and number of terminals per non-terminal.

The distributions were compared using mean, standard deviation, and information entropy. Table 1 shows some of these results.

Results

As can be seen in Figures 2 and 3, the distribution of the phenotypic search space is not evenly distributed. Also, the distributions can differ greatly by changing the gene length, wrapping, number of non-terminals, and number of terminals per non-terminal.

It is apparent that the grammar biases the phenotypic search space toward certain expressions. This could have a negative effect on the grammatical evolution algorithm, since, if the biased expressions are not functionally fit, they would clutter the search spaces.

```

S : : = <expr> (0)
<expr> : : = <expr> <op> <expr> (0)
          | ( <expr> <op> <expr> ) (1)
          | <var> ^ <const> (2)
          | <const> * <expr> (3)
<op> : : = + (0)
          | - (1)
          | * (2)
<var> : : = x (0)
          | y (1)
<const> : : = 1 (0)
           | 2 (1)
           | 3 (2)
           | 4 (3)
    
```

Figure 1: Basic polynomial grammar in BNF

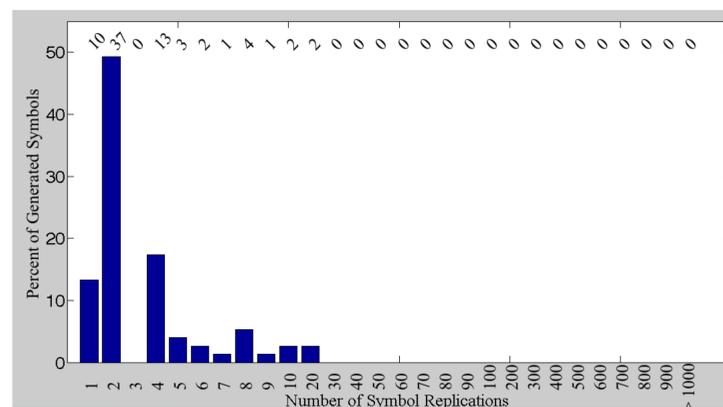


Figure 2: Polynomial grammar trial run with a gene length of 4 codons

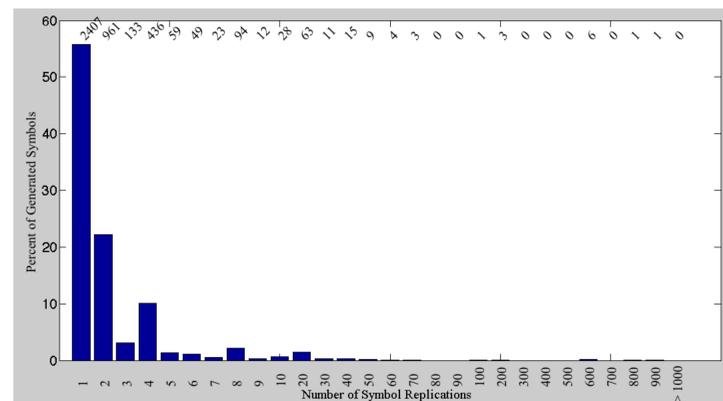


Figure 3: Polynomial grammar trial run with a gene length of 7

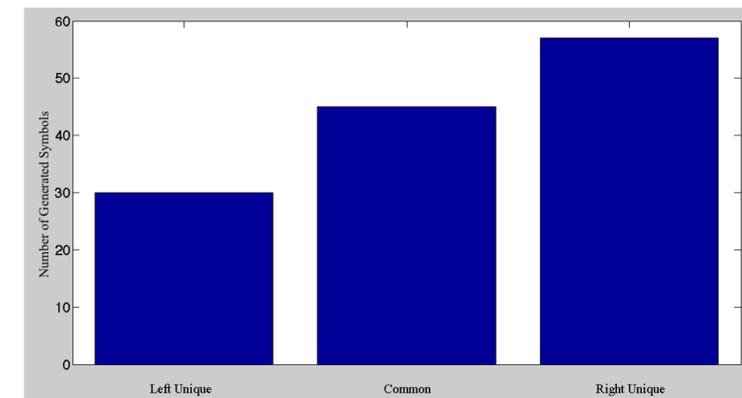


Figure 4: Comparison of Read from Left and Read from Right Grammars

Table 1: Statistics for Trials

Left or Right	Gene Length	Wrapping	Mean	Standard Deviation	Entropy	Phenotypes	Genotypes	Percent
L	4	1	3.41	2.68	5.69	75	256	29.30
L	5	1	3.52	6.18	7.21	291	1024	28.42
L	6	1	6.95	17.38	7.94	589	4096	14.38
L	4	2	3.46	4.58	5.53	74	256	28.91
L	5	2	2.69	7.36	7.11	380	1024	37.11
L	6	2	5.70	17.21	7.88	718	4096	17.53
L	7	2	3.79	27.66	9.50	4319	16384	26.36
R	4	1	2.51	2.15	6.26	102	256	39.84
R	5	1	2.73	5.22	7.58	375	1024	36.62
R	6	1	5.07	14.75	8.19	808	4096	19.73

Conclusion

The assumption that the distribution of the phenotypic search is evenly distributed is false for the typical grammars tested, and likely false for most grammars. A new method using MATLAB has been developed to study degeneracy on a more quantitative level. Many grammars, especially the grammars involved with symbolic regression, could be studied using the techniques developed. Further study of the effects on grammar variation could lead to improvements in writing grammars, which could lead to improvements in grammatical evolution's results and speed.

Acknowledgements: This work was made possible by the National Science Foundation REU Site: Bringing Us Together, Improving Communications and Lives EEC-1359476 and by Office of Naval Research (ONR) grant N00014-12-1-0347. Thanks to Professor of English Dr. Alfred Boysen for his critique in writing and speaking, and a special thanks to all of the faculty and students in the Electrical and Computer Engineering department and the Controls Lab for their help.