Examining Grammars and Grammatical Evolution in Dynamic Environments

Eoin Murphy

Natural Computing Research and Applications Group,
Univeristy College Dublin, Ireland.
eoin.murphy@ucd.ie

Abstract

This paper is concerned with the effect of the grammar type on grammatical evolution when evolving in dynamic environments. Both representation and dynamic environments have been recognised as important open issues in the field of genetic programming. This paper outlines the need for further study on both topics in the context of grammatical evolution, suggesting further inspiration be taken from nature in an attempt to improve the representations available to grammatical evolution. The research undertaken to date is listed, along with the future work to be completed.

1 Introduction

The goal of the forthcoming thesis is to explore the effect of the grammar type on Grammatical Evolution (GE) [15, 2] when evolving in a Dynamic Environment (DE).

The traditional Genetic Programming (GP) expression tree representation, popularised by Koza [10], is restricted by the property of closure to only support the use of a single data type. Grammar-Based GP (GBGP), an extension of standard GP, has the ability to ensure closure while allowing multiple types, as well as having a number of other benefits over standard GP.

GE, a linearised GBGP system which uses a genotype phenotype mapping and is one of the most widely applied GP methods today [11], traditionally uses a context-free grammar (CFG). The grammar is an integral component of the GE algorithm. This research aims to further improve the grammar-based representations available to GE.

One possible way is to strengthen the representation's analogy with nature. Evolutionary computing takes principles from the neo-Darwinian theory of natural selection, mimicking, in silica, these principles in an attempt to exploit them to solve problems. Genetic algorithms (GAs) and GP take further inspiration from nature in the form of *virtual DNA*, imitating one of nature's main representations, with GE taking one step further, making use of the genotype-phenotype mapping.

In addition, the majority of GP research to date has been on solving static problems. Little work has been done on dynamic problems, where the goal, the constraints or the environment of the problem can change with respect to time. Such problems are important as most real world problems inhabit DEs, including natural selection. DEs have been recognized as an open issue in the field [16, 2].

This research will focus on designing and examining the behaviour of different types of grammar-based representations when used with GE in DEs. The research undertaken to date has appeared as a number of publications [14, 13, 12]. The remainder of the paper is structured as follows: background in Sec. 2, research questions in Sec. 3, followed by work undertaken in Sec. 4 and future work in Sec. 5.

2 Background

The use of grammars in GP [11] has been shown to offer many advantages, e.g., the limitation imposed upon tree-based GP by closure can be avoided since type/syntactic information can be embedded in the grammar, allowing the genetic operations to operate on multi-type expressions. Grammars also allow the easy integration of domain knowledge into the search.

One of the main advantages of using GBGP is that the grammar allows the search space to be easily transformed. Grammars define the legal statements of the language, as well as the structures to be used to generate the solutions. By designing the grammar correctly, the search space can be greatly reduced, speeding up search. While speeding up search is desirable, in an extreme case the search space can be too heavily constrained, excluding the solution from the language and making the problem impossible to solve. This is similar to sufficiency in tree-based GP.

In addition to the advantages listed here, a survey of GBGP was completed by McKay et al. [11] listing further advantages, as well as costs, attributed with the use of grammars. Certain GP representations such as GE and TAG3P [4, 5] manage to avoid some of the disadvantages listed, such as alleviating the difficulty in creating new operators due to feasibility constraints.

The grammar traditionally used by GE, and indeed by many GBGP systems, is the CFG. While most work undertaken with GE makes use of regular CFGs to generate solutions, another key advantage of GBGP systems is how extensible they are, allowing many different grammar types to be explored. A comprehensive background can be found in the survey by McKay et al. [11], and Hemberg [3] completed a thesis on the use of grammars in GE.

Tree-adjunct grammars (TAGs) are a tree generating system. Unlike CFGs, TAGs can also generate some context-sensitive languages [8]. While TAGs originated in the field of natural language processing, they have been successful in GP, in the form of TAG3P [4, 5].

TAGs have some interesting properties making them ideal for GBGP. One of the most interesting properties is that TAGs overcome the disadvantage suffered by other GBGP systems, of having to deal with invalid individuals. Since the elementary trees used to compose TAG derivation trees are complete, the phenotype is always executable.

As mentioned in Sec. 1, strengthening the representation's analogy with nature could be beneficial. It has been shown that systems which allow the development of individuals to be effected by the external environment, as seen in nature, can greatly improve their performance, and hence speed up evolution [6]. The complete nature of TAG trees throughout derivation/development enables this notion to be explored in the context of GP/GE. Indeed, some work has already been completed on this topic by Hoang et al. [6].

In regards to DEs, little work has been undertaken with GE, and indeed GP, on the topic [2, 16]. However, it has been shown that in certain cases, varying environments can help speed up evolution [9]. Additionally, it has been theorised that the dynamism of nature has helped natural organisms evolve into such complex systems [9, 17, 18]. In fact, many real world problems inhabit DEs and if GP is to be applied to these fields, further research must be undertaken. As such, it is important to study DEs in order to understand how their properties effect natural evolution and how they can be exploited when evolving in silica.

The primary aim of this research is to discover new grammar types and representations which will help improve search in GE. Both GP and GE take inspiration from natural systems, further examination of these natural systems and their properties may be a good initial approach to creating new representations. This research will also attempt to better understand and improve GE's performance in DEs, in particular, this study aims to understand what makes a good representation for search in DEs.

3 Research Questions

In order to fulfill the research aims outlined directly above, the following questions need to be addressed:

- **Can further inspiration be taken from nature to help improve the representations used in GP?** One of the properties of GE is that it improves upon the analogy between GP and nature. By doing so it hopes to better mimic nature and improve GE's performance as an optimisation search algorithm. By committing further research into the field of representation and by trying to help strengthen this analogy even more, it is hoped to improve GE's current performance. One example of where the current analogy breaks down is that the CFG representation produces static, non-growing fully developed individuals, generated solely from the genotype.

- **The choice of representation can benefit performance in static environments. Do these benefits translate into DEs?** There has previously been studies on different grammar types and their effect on the GE algorithm [3]. What needs to be addressed is whether the different advantages and disadvantages of these representations remain when applied to DEs, and if so to what extent does the frequency or magnitude of change affect them. That is to say, what makes a good representation for DEs?

- **What are the properties of the representation which produce this performance increase? Can they be exploited?** Different grammar types and representations have different beneficial properties. New representations or GP systems could be developed to take advantage of these, perhaps even combining properties from multiple grammars to maximise performance.

4 Work Achieved to Date

A good deal of work has already been undertaken addressing the questions above, including three first author publications and another separate co-authored publication related to the research aims. This section outlines the work presented to date, the conclusions drawn from each body of work and how they address the research questions presented in Sec. 3.

4.1 Tree-Adjunct Grammatical Evolution

Tree-adjunct grammatical evolution (TAGE) [14] is an extension of the GE algorithm to work with TAGs. A comparison of the performance of TAGE against standard GE (using CFGs) across a range of benchmark problems drawn from GP literature was performed.

This study demonstrates that the use of TAGs in GE has a beneficial effect on GE's ability to move through the solution search space and to find successful solutions, see Fig. 1. TAGs have been found to have a positive effect on the performance of GP [4, 5], results which are shown by this study to carry over into the field of GE. The study showed that for four out of the five problems examined, TAGE found better solutions in fewer generations, as well as finding more perfect solutions than GE. Other interesting properties of using TAGs in GE were noted also, such as minimal growth of the chromosome, and the increase of connectivity in the search space which helps TAGE maintain diversity.

The decision to use TAGs was due to their developmental property which is synonymous to nature. This arises from the property of TAG derivation trees always being complete, unlike CFG derivation trees, allowing evaluation at different stages of development, a property that is synonymous to the development of embryos into adult organisms in nature. Natural development is not purely guided by genetics, but merely sets the blueprint for development, with external or environmental factors influencing development at each step. While this

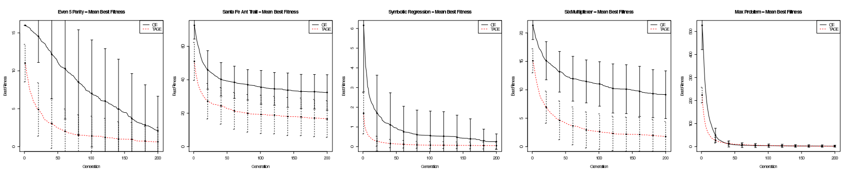

Figure 1: A comparison of GE and TAGE: Mean best fitness plots (minimising) across 100 runs at each generation with error bars of one standard deviation.

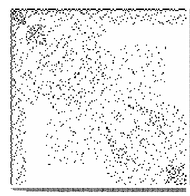

Figure 2: A heatmap where a black cell denotes phenotypes with neighbouring genotypes.

property was not taken advantage of directly in this study, it was shown that the TAG representation by itself improved the performance of GE.

4.2 Examining Mutation Landscapes In Grammar-Based GP

This study extends the previous by examining one of the properties of the TAG representation, namely the connectivity of the mutation landscape for a number of problems. The study compares the single Integer Flip Mutation (IFM) landscapes of GE and TAGE for a series of problems in an attempt to further understand how the change in representation effects each algorithm's ability to search [13].

Landscapes are a tool to help understand complex processes [7]. They have been employed here in an attempt to further understand how the use of TAGs in GE effects performance. Viewing the entire search space/landscape is difficult due to its large size and high complexity. To alleviate this problem, this study employs a novel method of visualisation little used in the field of GP, heat maps (an example can be seen in Fig. 2.

For the problems and grammars used in this study, it was found that phenotypes in the TAGE landscapes have a much higher degree of connectivity with the rest of the phenotypes than their counterparts in the GE landscapes. This may help explain the increased diversity within TAGE populations observed previously. Moreover, it was discovered that the connectivity in TAGE landscapes is much more evenly distributed between the other phenotypes in the landscape. Whereas in GE landscapes, shorter phenotypes are much more densely connected not only between themselves, but also, to a lesser extent, to the rest of the landscape.

Diversity is an important property for a representation to have when applied to DEs [2, 19]. The fact that TAGE landscapes have much higher connectivity than those of standard GE is encouraging and could benefit TAGE to maintain better diversity and to follow changing optima quicker.

4.3 A Comparison of GE and TAGE in DEs

With the previous two works in mind, this study performs a comparison of GE and TAGE when operating in DEs [12]. The definition of dynamism taken here is one where there is a functional change over time, i.e. the objective function changes with respect to time. By that definition the objective of evolution depends on time t:

$$f(x) := f(x, t) \tag{1}$$

where f is the fitness function, x is an individual and t is the current generation. This is considered as the simplest case and is one of the most commonly considered definitions in the evolutionary algorithms literature [19]. In order to compare the two setups a number of well understood static problems have been extended into the dynamic domain, enabling the form of the problem to change over time.

For the problems and grammars examined across the different setups, it was shown that TAGE on average performs slightly better than GE. However, on DEs with high magnitude of change, there is no clear advantage. It was also shown that the TAGE representation may help maintain a greater population memory than GE.

One clear result of this study is the indication that there is a need for the development of benchmark dynamic problems for GP, such as the moving peaks problem for GAs [1].

The study compares two different grammar types, one was chosen as a result of having a property analogous with what happens in nature (although not exploited in this case) and that the comparison took place while operating in a number of different DEs. It was shown that representation can have an effect on performance while operating in DEs, it was also shown that the benefits that were seen in the static environment were not directly observed in the dynamic domain.

5 Future Work

Expanding upon the work described in Sec. 4, the following research needs to be completed in order to properly address the aims of this forthcoming thesis.

Further work on the TAG representation for GE will be undertaken. While it has been shown that there are advantages to using TAGs, much work has been done using CFGs since GE's inception, some of which needs to be examined with TAGs. This includes work done on population initialisation. Initialisers will be created for TAG allowing similar structures to be generated as with

CFGs, enabling better comparisons between grammar types, and in doing so help further understand how the TAG representation operates.

In addition, the TAG representation has the inherent property of using the entire chromosome when mapping the phenotype, unlike standard GE which uses some percentage of codons. The result of which is that the only standard GP operator that can affect the size of a TAGE genotype, and hence phenotype, is crossover, which can be very destructive. This can be detrimental when the size of the ideal solution is far from the distribution of sizes in the population. With this in mind, other operators will be studied for use with TAGs in GE to try to alleviate this limitation.

One interesting property of the TAG representation is its ability to be evaluated at each stage of derivation, allowing the environment to influence the development of an individual as seen in nature. Building upon work by Hoang et al. [6], research will be done on examining the (dis)advantages of this property when used with GE. In particular, how a DE during development can effect performance [9].

In parallel to this, further research will be done into the natural computing and biological literature in an effort to identify other methods/properties of natural systems which might help create new, or improve current representations.

With regards to dynamic problems, as mentioned in Sec. 2, some work has been done on examining GE's performance in DEs, but there is a need for further work. A number of different approaches to search in DEs are outlined by Dempsey et al. [2] that need to studied in the context of GE, including memory, diversity, multiple populations, problem decomposition and evolvability. By doing so, better comparisons can be made on the performance of representations in DEs.

A big problem with the current state of the field is that there are no clearly defined benchmark problems for GP similar to the moving peaks problem for GAs [1]. In order to study GE's performance and behaviour in DEs, a small set of benchmark problems will be assembled. In order to achieve this, definitions will also be created to define what exactly is meant by a *DE*. A taxonomy of problems will be created, ranging in dynamism from small predictable change to large random change, in terms of the frequency and magnitude.

In conjunction with this, further study will be done on the role of the grammar type with regards to both the properties of dynamism above, frequency etc, as well as the approaches to solving dynamic problems listed by Dempsey et al. [2].

Acknowledgments

This research is based upon works supported by the Science Foundation Ireland under Grant No. 08/IN.1/I1868.

References

[1] J. Branke. Memory enhanced evolutionary algorithms for changing optimization problems. In *CEC 99. Proc. of the 1999 Congress on Evolutionary Computation, 1999*, volume 3, pages 1875–1882, 1999.

[2] I. Dempsey, M. O'Neill, and A. Brabazon. *Foundations in Grammatical Evolution for Dynamic Environments*. Studies in Computational Intelligence. Springer, 2009.

[3] E. Hemberg. *An Exploration of Grammars in Grammatical Evolution*. PhD thesis, University College Dublin, 2010.

[4] N. Hoai. *A Flexible Representation for Genetic Programming: Lessons from Natural Language Processing*. PhD thesis, University of New South Wales, 2004.

[5] N. Hoai, R. McKay, and D. Essam. Representation and structural difficulty in genetic programming. *IEEE Transactions on Evolutionary Computation*, 10(2):157–166, April 2006.

[6] T.-H. Hoang, D. Essam, R. I. B. McKay, and N. X. Hoai. Developmental evaluation in genetic programming: The TAG-based frame work. *International Journal of Knowledge-Based and Intelligent Engineering Systems*, 12(1):69–82, 2008.

[7] T. Jones. *Evolutionary Algorithms, Fitness Landscapes, and Search*. PhD thesis, University of New Mexico, 1995.

[8] A. Joshi and Y. Schabes. Tree-Adjoining Grammars. *Handbook of Formal Languages, Beyond Words*, 1997.

[9] N. Kashtan, E. Noor, and U. Alon. Varying environments can speed up evolution. *Proc. of the National Academy of Sciences*, 104(34), 2007.

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

[11] R. McKay, N. Hoai, P. Whigham, Y. Shan, and M. O'Neill. Grammar-based genetic programming: a survey. *Genetic Programming and Evolvable Machines*, 11(3-4):365–396, Sept. 2010.

[12] E. Murphy, M. O'Neill, and A. Brabazon. A comparison of ge and tage in dynamic environments. In *Proc. of the 13th Annual conference on Genetic and evolutionary computation*, GECCO '11, NY, USA, 2011. ACM.

[13] E. Murphy, M. O'Neill, and A. Brabazon. Examining mutation landscapes in grammar based genetic programming. In *Proc. of the 14th European Conference on Genetic Programming, EuroGP 2011*, volume 6621 of *LNCS*, pages 131–142, Turin, Italy, 27-29 Apr. 2011. Springer Verlag.

[14] E. Murphy, M. O'Neill, E. Galvan-Lopez, and A. Brabazon. Tree-adjunct grammatical evolution. In *2010 IEEE World Congress on Computational Intelligence*, pages 4449–4456, Barcelona, Spain, 18-23 July 2010. IEEE Press.

[15] M. O'Neill and C. Ryan. *Grammatical Evolution: Evolutionary Automatic Programming in an Arbitrary Language*. Kluwer Academic Publishers, USA, 2003.

[16] M. O'Neill, L. Vanneschi, S. Gustafson, and W. Banzhaf. Open issues in genetic programming. *Genetic Programming and Evolvable Machines*, 11(3-4):339–363, Sept. 2010.

[17] M. Parter, N. Kashtan, and U. Alon. Environmental variability and modularity of bacterial metabolic networks. *BMC Evolutionary Biology*, 7(1):169, 2007.

[18] M. Parter, N. Kashtan, and U. Alon. Facilitated variation: How evolution learns from past environments to generalize to new environments. *PLoS Comput Biol*, 4(11):e1000206, 11 2008.

[19] P. Rohlfshagen, P. K. Lehre, and X. Yao. Dynamic evolutionary optimisation: an analysis of frequency and magnitude of change. In *Proc. of the 11th Annual conference on Genetic and evolutionary computation*, GECCO '09, pages 1713–1720, USA, 2009. ACM.