Chapter 9
Conclusions and the Future

9.1 Summary

This book began by posing three questions concerning the application of GE to dynamic environments.

i. Is the correct infrastructure in place for GE to navigate dynamic environments? This required the investigation of the potential strengths inherent in GE and areas that require further attention for the effective application of GE to these environments
ii. Is GE capable of discovering new solutions when change occurs in the environment? Fundamental to the navigation of dynamic environments is that a population of solutions be capable of quickly transitioning to new areas of the solution landscape.
iii. Can GE maintain a diverse population of robust solutions capable of handling dynamic data? A major criticism of GA/GP approaches for dynamic environments is their tendency to converge. Maintaining diverse robust solutions is imperative for the successful application of GE to dynamic environments.

Combined with these, questions were posed relating to wider EC issues on the roles of memory and diversity in dynamic environments, along with how EC paradigms might best be tested in order to investigate these questions.

Following an introduction to Grammatical Evolution, a comprehensive survey of the state of the art for EC in dynamic environments was conducted. Stemming from this survey, a clear picture of the types of dynamic environments possible was outlined, along with the unifying of a number of researchers' work that identified the types of change that can occur in such environments. The various approaches researchers have adopted to navigate dynamic environments and overcome the extra issues specific to dynamic problems, were identified. Included in this was a discussion of performance measurement in the context of dynamic environments. An analysis of prior research was developed, producing a list of open research gaps.

I. Dempsey et al.: Foundations in Gramm. Evol. for Dyn. Environ., SCI 194, pp. 163–169.
springerlink.com

We then set out by providing an examination of GE in the context of dynamic environments, and reported our findings in some early experiments adopting non-stationary symbolic regression instances. Out of this, key strengths and areas for improvement were identified. Critical to the successful evolution of solutions for dynamic environments is an efficient mechanism for the generation and adaptation of constants, a mechanism that was lacking in the current state of the art in GE. A number of potential strengths were also identified, along with benefits unique to GE brought about through its use of BNF grammars.

Two novel constant generations mechanisms were introduced for GE or GP-type paradigms. The study then embarked on a series of experiments aimed at investigating the different methods of constant generation and adaptation. These experiments sought to focus on the specific issue of constant creation and adaptation in isolation. Both static and dynamic problems were examined with three different types of constant generation methods, along with experiments where the evolutionary search itself was used to determine the best method. This lead to the identification of Digit Concatenation with an ability to form expressions as the most efficient mechanism. Digit Concatenation presented an evolvable representation at the grammar level that provided advantages such as greater accuracy and incremental evolution, over the other methods explored.

Further exploration of the benefits of BNF grammars was then conducted with the introduction of another novel method of constant generations through the use of meta-Grammars and $(GE)^2$. Experiments were executed with the $(GE)^2$ paradigm using a population half the size of that used in prior experiments, so as to maintain an equal computational effort. Despite the reduced population size and an increase of the search space, brought about through the solution-grammar chromosome, $(GE)^2$ performed competitively with standard GE, while also demonstrating an increased rate of evolution. For dynamic experiments, $(GE)^2$ was seen to adapt at a greater rate over some prior experiments showing significant potential for further exploration.

Having completed this, experiments were then executed aimed at gaining a wider understanding of the behaviour of GE in the complex setting of a real-world problem, that of trading on financial markets. A background to modern financial theory was first presented to provide a backdrop to the significance and complexity of the domain. This was followed with a survey of prior work carried out in applying EC techniques to the domain. Stemming from this, a number of controlled artificial static experiments were conducted. This approach was adopted in order to gain a clearer understanding of the system's behaviour under a noise-free setup. These experiments demonstrated the ability of GE to perform in the domain and also highlighted weaknesses of the static approach.

This then provided the scope to apply GE to dynamic financial data sets. A variety of experiments were conducted that sought to build up complexity, allowing a greater understanding of GE's behaviour to be developed at each level. First GE was seen to trade more conservatively than random trading for shuffled data sets. Progressing onto the normal unshuffled data, the adaptive approach yielded superior returns over random trading. Finally in comparison to a static approach, the adaptive approach also produced superior results. In addition these experiments also served to highlight GE's ability to maintain diversity within the population as a function of the potential fitnesses in the search space.

Arising from this book, these series of experiments provide evidence for the benefits of novel constant creation methods. It identifies evolvability and genotype-to-phenotype mapping as being key assets for evolutionary algorithms applied to dynamic environments, circumventing issues surrounding the fundamental problem in the literature thus far, that of convergence and reduction in diversity.

9.2 Research Results

Through the gathering together of the various literature on the topic of EC in dynamic environments and the subsequent execution of a series of experiments in the domain, a number of key insights are provided on different levels. This section will review the research results of this book.

9.2.1 Analysis of GE in Dynamic Environments

This book for the first time analysed GE in a dynamic environment. The issues surrounding evolution in dynamic environments are different to those of a static one. In Chapter 4, key potential strengths of GE for dynamic environments, along with areas where it could be improved were identified and analysed alongside work in other areas of EC. Subsequent chapters then introduced effective evolvable methods for constant generation and adaptation tailored for dynamic problems; examined the diversity-maintaining effects of the genotype-to-phenotype mapping; analysed $(\mathrm{GE})^2$ in dynamic environments; and demonstrated the effectiveness of GE in dynamic environments. Combined with this, as far as the authors are aware, this book represents the first comprehensive investigation of a GP-type paradigm set in a dynamic environment.

9.2.2 Comprehensive Literature Review for EC in Dynamic Environments

Chapter 3 provides a detailed survey of the various facets and issues surrounding the application of EC techniques to dynamic problems. The chapter began by questioning what constituted a dynamic environment and what the possible types of change were that could occur. This led to the unifying of a number of author's works to form a new taxonomy that then provided a foundation for analysis of the various approaches researchers have adopted in extending EC to handle dynamic environments. Comprehensive coverage was given of the various approaches taken by researchers. This was followed by an examination of the different metrics adopted to measure performance over dynamic landscapes as well as the types of benchmarks used in evaluating approaches. This chapter uncovered a number of gaps in the state of the art, where very little attention is paid to GP-type paradigms in dynamic environments, and too much attention is paid to simplistic benchmarks and equipping algorithms with forms of explicit memory.

9.2.3 Extensions to GE's Ability to Create and Adapt Constants

The majority of phenotypic solutions generated by evolutionary algorithms contain constants. Where the environment is dynamic, there is then a requirement to generate new and/or adapt existing constants. Thus, an effective and efficient mechanism for creating and adapting constants is imperative. GE and GP generally adopt an inflexible approach to the creation and adaptation of constants. This book explored the properties and flexibility of BNF grammars, unique to GE, to create efficient and flexible constant generation and adaptation mechanisms. Issues such as ease of evolution and accuracy were uncovered and addressed. Digit Concatenation with the ability to form expressions was identified as being the best method for constant creation and adaptation.

9.2.4 Novel Methods for Constant Creation in GE

In the process of investigating and identifying the best method for constant creation, two novel methods for constant creation were introduced: Digit Concatenation, Persistant Random constants, and constant generation through $(GE)^2$. Digit Concatenation provides an ability to continuously create new constants through evolution and incrementally improve the fitness of solutions by adapting individual constants at the digit level. Persistant Random Constants presents a variation on GP's ERC. ERC, the widely adopted mechanism for constant creation in GP, suffers from a number of flaws as outlined in Chapter 5. For Persistant Random Constants, similar to ERC, a large population of random constants is initially generated. However, unlike ERC these constants remain available for evolution to reintroduce to the population at any point. This facilitates the maintenance of a more diverse selection of constants. Constant creation through $(GE)^2$ using the Digit Concatenation method as a base in the meta-Grammar allowed the automatic biasing by the solution-grammar chromosome of useful constants.

9.2.5 Identification of Diversity Being a Function of Potential in the Environment

If application of EC to dynamic environments is to be successful, the legacy problem of static environments, where a population is encouraged to converge to an optimal solution, must be overcome. For dynamic environments, the maintenance of diversity is key. This diversity allows the population to provide wide coverage of the solution landscape as it changes over time. GE's genotype-to-phenotype mapping is an ideal solution to this problem, where in the genotype space, individuals can be dispersed across a wide spectrum with variable lengths and different genotypic values. These genotypes are then mapped into the phenotypic solution space. In Chapter 8, it was found that this mechanism equips GE with an ability to avoid convergence and furthermore, enables it to maintain phenotypic diversity in line with opportunities presented on the fitness landscape. Phenotypic diversity becomes a function of the fitness landscape, while the genotype is free to evolve along neutral networks.

9.2.6 Identification of Two Levels of Evolvability in GE

This book encountered the benefits of evolvability in two chapters. A high level of evolvability imbues the ability upon a population to incrementally evolve to better solutions, when it is in a local optima navigating to the global optimum or following an optima across a changing landscape. At the level of the BNF grammar in Chapter 5, Digit Concatenation was seen to achieve more accurate fitnesses because of its ability to incrementally evolve to the best solution. In Chapter 8, the population of solutions was seen to quickly adapt to new solutions after periods where the phenotypic diversity decreased due to the index making a loss. This speedy adaptation was brought about through GE's genotype-to-phenotype mapping, allowing it to continue evolutionary search through neutral mutations, while still maintaining phenotypic fitness.

9.2.7 Experimental Evidence of the Evolution of Robust Solutions over Dynamic Data

In Chapter 7, the weakness of evolving a population over a static data set was highlighted, where optimal solutions were produced but these solutions were brittle and overfit the training data. Chapter 8 provided evidence that where a population was progressively trained over dynamic data, the resulting solutions were more robust. On average in 50% of window increments the best-performing solution survived to trade again on the live data. Through being exposed to this dynamic data, GE was able to uncover more valuable underlying data that enabled it to produce solutions that transferred successfully to live or out-of-sample data.

9.2.8 *Experimental Evidence of the Presence of Useful Information in Real-World Financial Historical Time Series*

This book simulated live trading over historical financial time series. As a means of generating trading rules, a relatively simple grammar was used that included one technical indicator, the moving average indicator. However, when combined with GE, this grammar was able to produce rules that outperformed a benchmark buy-and-hold strategy in 23% of runs conducted on the Nikkei 225 data, while also producing competitive results on a strongly upward trending S&P 500 index. This implies for the Nikkei 225 that, for the period under investigation, the index did not exhibit efficient price formation when compared against the buy-and-hold benchmark, as the index was beaten through exclusive use of technical analysis on publicly available information. However, when considering the EMH as a joint hypothesis it becomes difficult to reject efficiency as the chosen benchmark may not be suitable. the use of financial data represents the introduction of a new real-world test-bed for the analysis of EC approaches to dynamic environments that experience complex types of change. Prior benchmarks focus on Markov or Deterministic types of change and, unlike these benchmarks, no standardised benchmark code is required, as the driver is widely available standardised data.

9.3 Opportunities for Future Research

This study represents an initial step in the exploration and application of GE to dynamic environments and as such represents a foundation for further application and development of GE in this area.

Chapter 3 outlined a number of approaches other researchers had adopted in dealing with dynamic environments. Memory, multiple populations, and problem decomposition were not explored in this book. Of these, there exists much scope for the exploration of the existence of memory within meta-Grammars and $(GE)^2$. For deterministic types of change, it may be possible for the solution-grammar chromosome to evolve useful biases and terminal subsets, allowing the solution chromosome to evolve a representation that allows efficient switching between solutions.

Further research opportunities in this area also exist in examining the potential of the solution-grammar chromosomes to evolve useful building blocks, which ties in with the idea of forming a memory in this chromosome. Useful sections of code or ADFs may be evolved in the solution-grammar chromosome, allowing the solution chromosome to then essentially conduct its evolutionary search with bigger pieces. Such an approach may enable $(GE)^2$ to remember *how* to produce fit solutions, not just fit solutions themselves. However, attention must be paid to maintaining levels of evolvability, which was demonstrated in this book to be of importance for dynamic environments.

A natural extension of this work is in replacing the standard GE mapping process with that of πGE. πGE has demonstrated an ability to produce superior results to those of standard GE in static problems. The application of πGE to dynamic problems may lead to an improvement in results of the same order.

This book conducted an examination of GE in the setting of a real-world problem that experienced complex change. Further application of EC paradigms to dynamic real-world problems may focus research on issues in those domains, rather than on limitations of catch-all benchmark problems. Exclusive research against benchmark problems runs the risk of tailoring algorithms to such problems.

9.4 Finally

Evolution in the natural world is an ongoing process that has led to a diverse array of complex living organisms that are capable of surviving in challenging dynamic environments. These organisms continue to evolve and survive as their environment changes or else fail to evolve and face extinction. As researchers, the utilisation of this process presents immense promise in providing dynamic solutions to the many real-world problems that experience some form of change with time. This book has provided a foundation on the road to realising that goal. The types of changes that can be faced by modellers were described. GE was then analysed in this context developing a greater understanding of evolvability, diversity and the quality of solutions evolved on dynamic data. This area represents a great opportunity for future research that must be conducted if the full potential of the evolutionary process is to be realised in the real-world, and this book represents just the initial foundations upon which future research can take place.