

Ecosystem Curation in Genetic Improvement for Emergent Software Systems

Zsolt Németh
School of Computing and
Communications, Lancaster
University, UK
UK
z.nemeth@lancaster.ac.uk

Penn Faulkner Rainford
Department of Computer Science,
University of York, UK
UK
penn.rainford@york.ac.uk

Barry Porter
School of Computing and
Communications, Lancaster
University, UK
UK
b.f.porter@lancaster.ac.uk

ABSTRACT

Emergent software systems are composed, and continuously re-composed at runtime, from a large pool of small potential building blocks with the aim of responding to changes in the deployment environment [3]. The approach assumes that each building block has a set of available variations in the selection pool, such that the most appropriate collection of variants can be composed according to the current conditions, with some objective function in mind. Populating such a pool of implementation variation, however, is not a trivial task, and existing work has examined the use of Genetic Improvement (GI) to drive this process [6].

1 INTRODUCTION

In existing GI-driven approaches in this space, because the building blocks being considered are very small and are improved in isolation, researchers have used a mixture of new code synthesis with traditional GI mutation/crossover operators to gain sufficient novel genetic material. To cope with the resulting search space size in a general-purpose programming language, research has suggested the use of phylogenetic analysis to help drive (rather than to explain) evolution; this examined the independent effects of both crossover and different mutation types [6], and was also used to demonstrate combined lineage selection and optimisation in a single run [5]. While this use of phylogenetics offers a level of automated search guidance, by itself it remains limited to single genetic pools. In this paper we propose an abstraction shift towards *ecosystem curation* as a top-level driver for GI processes in order to balance the need for novel genetic material (breadth of search) with exploitation of high-utility regions (depth of search).

2 CHALLENGES

In contrast to approaches where genetic improvement occurs in a single monolithic population, we will consider curated and guided ecosystems of both horizontally and vertically divided population groups, combined with phylogenetic-inspired reasoning along three axes: the simultaneous evolution, and controlled periodic interaction, of multiple distinct populations; the definition and curation of *species* within populations the guidance of evolutionary processes based on species cardinality; and the cooperative injection of highly inhomogeneous individuals at the inception of GI processes in ways that are mutually beneficial to a population and do not cause immediate dominance of a single genetic lineage. Taken together we will use this set of approaches to form ecosystems of genetic material, with the aim of balancing the navigation of breadth and depth, and

maintaining diversity to rapidly evolve new individuals towards novel deployment conditions.

(i) *Horizontally divided populations.* We will study the effects of *simultaneously evolving spatially separated genetic pools* that occasionally interact. Each such isolated pool has its own dynamicity, direction of evolution, potentially different parameters and training sets, and from time to time some individuals may cross the boundaries. Gene transfer in this case realises crossover between populations instead of individuals.

In a single genetic pool – after the first few random steps – the direction of evolution is governed by the fittest individuals. Hence, the fitness landscape is not explored in breadth but along a few directions of depth, likely leading to some local optima. In multiple genetic pools these exploration paths are multiplied and by meta-crossover we expect the potential to steer away from local optima and open new paths for exploration towards better individuals.

We will investigate metrics for comparing populations and strategies, possibly involving further nature metaphors, to govern the selection and transfer of individuals across islands. We will examine whether migrating individuals blend into the receiving population or bring new traits and change the direction of evolution, and if such migration can be constructively directed towards improved results. We will track how the overall utility of populations and fitness of individuals change across such crossovers.

(ii) *Vertically divided populations by species.* When we consider non-functional properties like performance as our improvement target, it is possible to distinguish individuals of a population into distinct species. For hash table implementations, their hash function can be characterised into species by a distance of actual distribution from a uniform distribution; a similar metric may be used for scheduling algorithms. This speciation effectively separates the algorithm from the way it is represented in source code.

Species can be used to characterise the meaningful diversity of a population that further can be quantified by diversity indices [7]. The neutral planes of the fitness landscape are made of individuals with different genotypes but with phenotypes that indicate the same species (in other words, most of these individuals are variants of essentially same algorithm). The notion of species could help to identify these neutral spaces and also steer the evolutionary process towards new species, an objective similar to novelty search [1]. The concept of species can be used to maintain a measurable degree of diversity, to ensure continuity of exploration towards more distant points of a search space, and potentially to help motivate leaps over neutral planes in the search for novel species.

The definition of species, as a metric, is crucial to its viability: it must be selective, characteristic, and insensitive to implementation details while being easily computable. The notion is application-specific, but our pilot application of hash functions suggests we can find suitable metrics. We intend to observe the creation, lifespan, decline and extinction of species and investigate what evolutionary processes lead to these events and how they relate to the overall population lifecycle. Our investigation will focus on relations between diversity, population utility and individual fitness and whether control of diversity has an effect on improvement.

(iii) *Highly inhomogeneous populations.* Another form of ecosystem curation is the study of co-existence and interaction between individuals from very different origin, meaning extreme variations in fitness and utility within one population. This is different from (ii) where species evolve from the same origin; here a vertical separation is deliberately introduced from the first generation. This offers multiple distant starting points and a profound diversity of the gene pool that promises a good exploration of the fitness landscape.

While this may appear simple, as a facet of ecosystem curation we are interested in avoiding the inevitable short-term dominance of a single species on the population which almost immediately removes diversity. We will investigate techniques which ensure transfer and upkeep of diverse genetic material, maintaining the utility of such highly inhomogeneous populations, while still allowing GI to push towards better individuals overall. One such approach is the definition of individuals that are immutable, and which carry and transfer genetic material but otherwise do not contribute to the ecosystem of the population. Another approach is the study of dynamic models, like the Lotka-Volterra predator-prey model [?], that capture the biological fact that the fittest individuals can survive in the sufficient presence of lesser fit ones only (where, for example, the predator-prey model ensures the survival of the prey).

(iv) *Curated genetic material.* Our fourth research direction builds on (i), (ii), and (iii), to incorporate their results into ecosystems of populations. The study of ecosystems is not autotelic beyond the scientific curiosity but has a practical goal: to deliberately curate and store high-utility genetic material for later combination or mutation. In the context of emergent software systems, which encounter novel operating environments necessitating new implementation variants, having a highly-evolvable gene pool may provide starting points for rapidly accommodating new environments.

To this end we envision the controlled breeding of genetic material composing the effects of separately evolving genetic interacting pools and genetic material of different origins. The aim here is to find ‘interesting’ points in the fitness landscape that are already specialised in some direction, yet have the potential for further evolution. We plan to experiment with hierarchical populations where individuals keep migrating according to their fitness and populations that are specialised to certain directions.

Ecosystems and multi-populations relate to research on co-evolution [8], but instead of cooperative [4] or competitive [2] strategies, we will curate the genetic makeup of ecosystems overall by studying phylogenetics and speciation, driving inter-population mixing alongside the curation of highly-evolvable material.

3 DISCUSSION

The curation of genetic ecosystems, through populations, species, and highly-evolvable individuals, presents the opportunity to combine search by breadth and depth, and to guide the level of ongoing genetic diversity available to a GI process.

We will create a system able to automatically capture and analyse phylogenetic data to understand changes in fitness between generations relative to the mutations and crossovers. We plan to surface further details of the effects of various genetic transformations on the fitness, utility, size and other attributes of the population. The study combining different phylogenetic histories in the context of antecedent transformations is essential to understand the interaction of separately evolved groups.

We will then combine these insights with the identification of species, and the mixing of populations, to understand and guide how GI overall is operating as a search process. At the same time we aim to maintain a suitable balance between external guidance of a GI process via ecosystem-level analysis, alongside the most desirable existing properties of GI-based search in locating good individuals through weighted stochastic gene combination. Finding the right balance to this equation is a key research question underpinning the whole of our proposed approach.

ACKNOWLEDGMENTS

This work was supported by the Leverhulme Trust Research Grant ‘Genetic Improvement for Emergent Software’, RPG-2022-109.

REFERENCES

- [1] Stephane Doncieux, Alban Laflaquière, and Alexandre Coninx. 2019. Novelty search: a theoretical perspective. In *Proceedings of the Genetic and Evolutionary Computation Conference (Prague, Czech Republic) (GECCO '19)*. Association for Computing Machinery, New York, NY, USA, 99–106.
- [2] Jan Paredis. 1995. Coevolutionary Computation. *Artificial Life* 2, 4 (07 1995), 355–375. <https://doi.org/10.1162/artl.1995.2.4.355>
- [3] Barry Porter, Matthew Grieves, Roberto Rodrigues Filho, and David Leslie. 2016. RE^X: A Development Platform and Online Learning Approach for Runtime Emergent Software Systems. In *Symposium on Operating Systems Design and Implementation*. USENIX, 333–348.
- [4] Mitchell A. Potter and Kenneth A. De Jong. 1994. A cooperative coevolutionary approach to function optimization. In *Parallel Problem Solving from Nature – PPSN III*, Yuval Davidor, Hans-Paul Schwefel, and Reinhard Männer (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 249–257.
- [5] Penny Faulkner Rainford and Barry Porter. 2022. Lineage Selection in Mixed Populations for Genetic Improvement. In *ALIFE 2022: The 2022 Conference on Artificial Life*. MIT Press. https://doi.org/10.1162/isal_a_00494
- [6] Penny Faulkner Rainford and Barry Porter. 2022. Using phylogenetic analysis to enhance genetic improvement. In *GECCO '22: Genetic and Evolutionary Computation Conference, Boston, Massachusetts, USA, July 9 - 13, 2022*, Jonathan E. Fieldsend and Markus Wagner (Eds.). ACM, 849–857. <https://doi.org/10.1145/3512290.3528789>
- [7] Michael Roswell, Jonathan Dushoff, and Rachael Winfree. 2021. A conceptual guide to measuring species diversity. *Oikos* 130, 3 (2021), 321–338. <https://doi.org/10.1111/oik.07202>
- [8] David R. White, Andrea Arcuri, and John A. Clark. 2011. Evolutionary Improvement of Programs. *IEEE Transactions on Evolutionary Computation* 15, 4 (2011), 515–538. <https://doi.org/10.1109/TEVC.2010.2083669>