Lexicase Selection and the Diversity of Quality

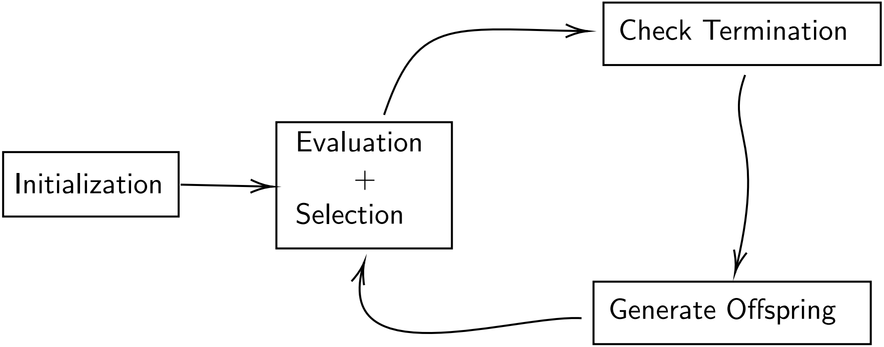
Ryan Boldi

University of Massachusetts Amherst

27th May 2022

UMassAmherst | Manning College of Information & Computer Sciences

General Evolutionary Algorithm



Evaluation + Selection

Generally, there are many things that we want to promote in a population. This could be performance on multiple tasks (ER) or test cases (GP).

Evaluation + Selection

Generally, there are many things that we want to promote in a population. This could be performance on multiple tasks (ER) or test cases (GP).

How do we pick parents that are good at these things?

Evaluation + Selection

Generally, there are many things that we want to promote in a population. This could be performance on multiple tasks (ER) or test cases (GP).

How do we pick parents that are good at these things?

- How do we measure the goodness of a parent?

Evaluation + Selection

Generally, there are many things that we want to promote in a population. This could be performance on multiple tasks (ER) or test cases (GP).

How do we pick parents that are good at these things?

- How do we measure the goodness of a parent?
- How do we compare parents that are good at different things?

Evaluation + Selection

Generally, there are many things that we want to promote in a population. This could be performance on multiple tasks (ER) or test cases (GP).

How do we pick parents that are good at these things?

- How do we measure the goodness of a parent?
- How do we compare parents that are good at different things?
- How many children should each parent get?

Answer 1: Aggregation

Take everything that matters to you, score each parent on each thing, and sum them together to get one fitness value.

Problems:

- Some things might be more important than others.

Answer 1: Aggregation

Take everything that matters to you, score each parent on each thing, and sum them together to get one fitness value.

Problems:

- Some things might be more important than others.
- Causes you to compromise on scores between tasks.

Answer 1: Aggregation

Take everything that matters to you, score each parent on each thing, and sum them together to get one fitness value.

Problems:

- Some things might be more important than others.
- Causes you to compromise on scores between tasks.
- Obscures information

Answer 1: Aggregation

Take everything that matters to you, score each parent on each thing, and sum them together to get one fitness value.

Problems:

- Some things might be more important than others.
- Causes you to compromise on scores between tasks.
- Obscures information
- Generalists instead of Specialists.

Answer 2: Multi-objective optimization

Maintain individuals on the Pareto front.

Answer 2: Multi-objective optimization

Maintain individuals on the Pareto front.
Problems:

- Curse of Dimensionality when number of objectives is high

Answer 2: Multi-objective optimization

Maintain individuals on the Pareto front.
Problems:

- Curse of Dimensionality when number of objectives is high
- Need to tune many hyperparameters

Answer 2: Multi-objective optimization

Maintain individuals on the Pareto front.
Problems:

- Curse of Dimensionality when number of objectives is high
- Need to tune many hyperparameters
- Sometimes still aggregate errors

Answer 3: Lexicase Selection

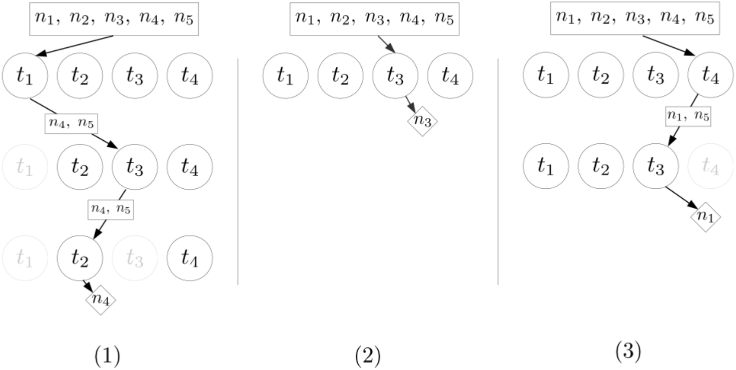
Selecting individuals based on a random ordering of cases.

Basic Algorithm:

1. Shuffle cases
2. Keep individual(s) that are elite on the first case.
3. If one individual remains, return it
4. Else, repeat with the next case in the order
5. If cases have run out, return a random individual

Lee Spector. "Assessment of problem modality by differential performance of lexicase selection in genetic programming: a preliminary report". In: *Proceedings of the 14th annual conference companion on Genetic and evolutionary computation*. 2012

Lexicase Selection



William La Cava, Thomas Helmuth, Lee Spector, and Jason H Moore. "A probabilistic and multi-objective analysis of lexicase selection and ε-lexicase selection". In: *Evolutionary Computation* 3 (2019)

Epsilon-Lexicase Selection

If we are operating in a domain with float fitness values, there are usually no ties for the "elite" individual. This means Lexicase selection will usually end up selecting individuals based on one test case.

To fix this, we relax the elite condition to now maintain individuals that are within ε of the elite individual.

This turns out to be a very powerful symbolic regression technique.

William La Cava, Lee Spector, and Kourosh Danai. "Epsilon-lexicase selection for regression". In: *Proceedings of the Genetic and Evolutionary Computation Conference 2016*. 2016

Who is Rewarded?

Lexicase selection promotes the selection of the following types of individuals:

- Good at things that others are not good at
- Good at difficult things
- Good at a *unique combination* of things $\rightarrow$ diversity of quality

Note the biological inspiration here.

Lexicase Selection in GP

Lexicase selection outperformed tournament selection and implicit fitness sharing across a variety of benchmark program synthesis problems.

Thomas Helmuth and Lee Spector. "General program synthesis benchmark suite". In: *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation.* 2015.

Problem	Tourn	IFS	Lex	Size
Number IO	68	72	<u>98</u>	5
Small Or Large	3	3	5	27
For Loop Index	0	0	1	21
Compare String Lengths	3	6	7	11
Double Letters	0	0	6	20
Collatz Numbers	0	0	0	
Replace Space with Newline	8	16	<u>51</u>	9
String Differences	0	0	0	
Even Squares	0	0	2	37
Wallis Pi	0	0	0	
String Lengths Backwards	7	10	<u>66</u>	9
Last Index of Zero	8	4	<u>21</u>	5
Vector Average	14	13	16	7
Count Odds	0	0	<u>8</u>	7
Mirror Image	46	64	<u>78</u>	4
Super Anagrams	0	0	0	
Sum of Squares	2	0	6	7
Vectors Summed	0	0	1	11
X-Word Lines	0	0	<u>8</u>	15
Pig Latin	0	0	0	
Negative To Zero	10	8	<u>45</u>	8
Scrabble Score	0	0	2	14
Word Stats	0	0	0	
Checksum	0	0	0	
Digits	0	1	7	20
Grade	0	0	4	52
Median	7	43	45	10
Smallest	75	<u>98</u>	81	8
Syllables	1	7	18	14
Problems Solved	13	13	22	

Lexicase Selection in Evolutionary Robotics

Jared M. Moore and Adam Stanton. "Lexicase selection outperforms previous strategies for incremental evolution of virtual creature controllers". In: *ECAL*. ed. by Carole Knibbe et al. MIT Press, 2017.

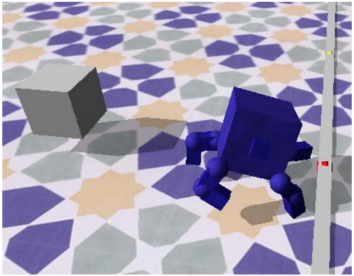
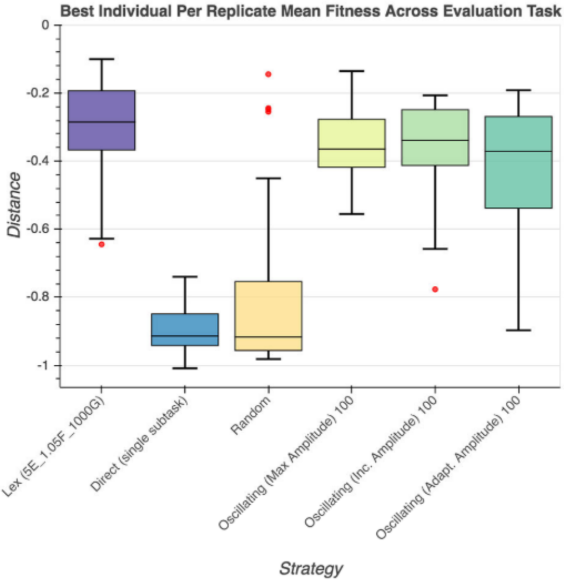


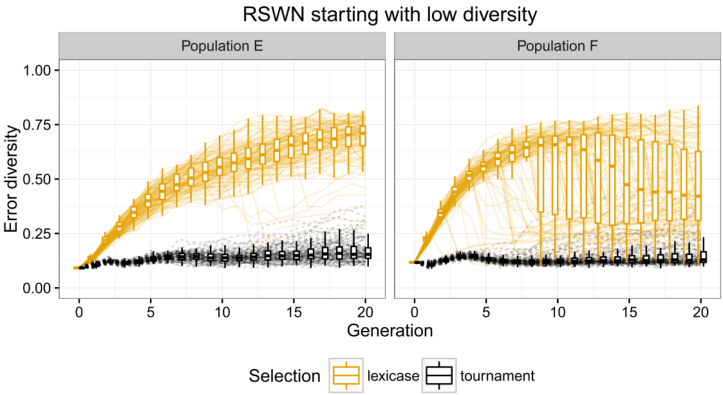
Figure 1: The quadrupedal animat and simulation environment in this study. The animat is tasked with crossing a wall (image right) and moving towards a target, represented by the box (image left).

Lexicase Selection in Evolutionary Robotics

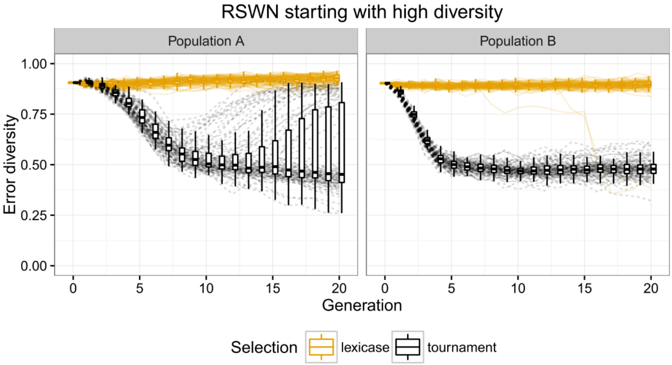


Promoting Diversity I

Lexicase selection has been found to promote diversity across a variety of fields without explicitly selecting for it.



Promoting Diversity II



Thomas Helmuth, Nicholas Freitag McPhee, and Lee Spector. "Effects of lexicase and tournament selection on diversity recovery and maintenance". In: *GECCO 2016 Companion - Proceedings of the 2016 Genetic and Evolutionary Computation Conference* (July 2016)

Lexicase Selection in More Fields

- Learning Classifier Systems
  Sneha Aenugu and Lee Spector. "Lexicase selection in learning classifier systems". In: *Proceedings of the Genetic and Evolutionary Computation Conference*. 2019

- Deep Learning
  Li Ding and Lee Spector. "Optimizing Neural Networks with Gradient Lexicase Selection". In: *International Conference on Learning Representations*. 2022

- Boolean constraint satisfaction with a GA
  Blossom Metevier, Anil Kumar Saini, and Lee Spector. "Lexicase Selection Beyond Genetic Programming". In: *Genetic Programming Theory and Practice XVI*. ed. by Wolfgang Banzhaf, Lee Spector, and Leigh Sheneman. Cham: Springer International Publishing, 2019

Multi-Objective Optimization

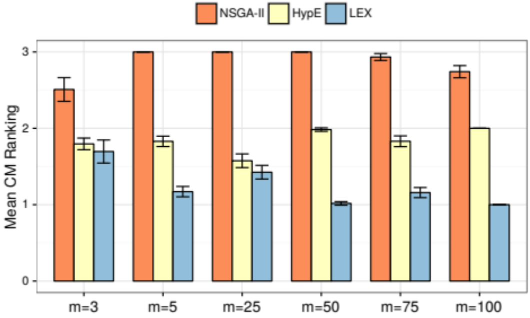
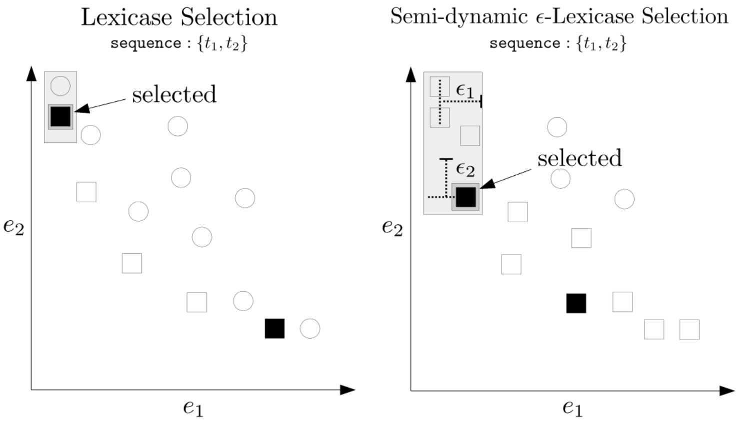


Figure 4: Average CM rankings of each algorithm as a function of m.

William La Cava and Jason H. Moore. "An Analysis of -Lexicase Selection for Large-Scale Many-Objective Optimization". In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. GECCO '18. Kyoto, Japan: Association for Computing Machinery, 2018

Selected Individuals are Pareto Optimal

(ε-Pareto Optimal for ε-Lexicase)



William La Cava, Thomas Helmuth, Lee Spector, and Jason H Moore. "A probabilistic and multi-objective analysis of lexicase selection and ε-lexicase selection". In: *Evolutionary Computation* 3 (2019)

Large-Scale Systems

What do we do when error computations are very expensive?

Li Ding, Ryan Boldi, Thomas Helmuth, and Lee Spector. "Lexicase Selection at Scale". In: *Genetic and Evolutionary Computation Conference Companion (GECCO '22 Companion), July 9–13, 2022, Boston, MA, USA*. 2022

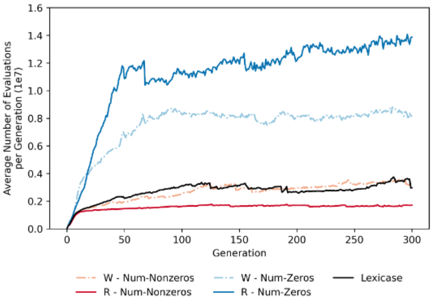


Figure 2: Average number of evaluations performed in a given active generation over evolutionary time while solving the *Last Index of Zeros* problem.

Down-sampling

Instead of using the entire training set every generation, use a subset of the cases (5-10%)

- Fewer individual evaluations every generation.
- Now, you can run for more generations for the same computational budget.
- Has been found to significantly improve success rates in GP runs

Jose Guadalupe Hernandez, Alexander Lalejini, Emily Dolson, and Charles Ofria. "Random subsampling improves performance in lexicase selection". In: *GECCO '19: Proceedings of the Genetic and Evolutionary Computation Conference Companion*. Prague, Czech Republic: ACM, 13-17 7 2019

Thomas Helmuth and Lee Spector. "Problem-solving benefits of down-sampled lexicase selection". In: *Artificial Life* (June 2021). ISSN: 1530-9185. arXiv: 2106.06085

Rolling Down-samples

Instead of randomly down-sampling every generation, what if there was some continuity between the "task-environments" every generation?

Does not seem to help for GP runs, but opens a new research direction.

Ryan Boldi, Thomas Helmuth, and Lee Spector. "Exploring Environmental Change for Down-Sampled Lexicase Selection". In: ALIFE 2022: The 2022 Conference on Artificial Life. July 2022

Future Work

Currently being written up:

- Changing the way we down-sample beyond purely random/rolling. $\rightarrow$ positive results so far.

Interesting things for future exploration:

- Lexicase Selection for Quality Diversity.
  - Elite/Not-Elite swapped for behavior present or not present.

QUESTIONS?