Analyzing the influence of Selection on Genetic Programming's Generalization ability in Symbolic Regression

A comparison of epsilon-lexicase Selection and Tournament Selection

2022-06-29

Introduction

Research Question

Experimental Study

Results

Conclusions

Limitations and open Questions

Introduction

Research Question

Research Question

- Does the usage of ϵ-lexicase parent selection influence the generalization behaviour of genetic programming in symbolic regression if compared to tournament selection?
-

Genetic Programming

- A metaheuristic that searches for computer programs that solve a given problem
- Inventor: John R. Koza[1]
- Evolutionary algorithm that simulates the process of Darwinian evolution:
  1. Population based
  2. The quality of solutions is evaluated by a fitness function
  3. Selection: Solutions are selected based on their individual fitness
  4. Variation: Mutation and recombination of solutions
- Unique Features:
  - Evolve solutions of variable length and structure
  - Solutions are typically represented by recursive tree structures

[1] Koza (1992)

Parent Selection

- Operator that selects individual solutions from the population for reproduction and mutation
- Most commonly used selection operator in GP: Tournament selection[2]
- Intuition: High chance for "generalist" solutions to be selected since it is based on aggregated fitness scores

[2] Fang and Li (2010), p.181

epsilon-Lexicase Selection

- Recent alternative: Lexicase Selection and it's variation ϵ-lexicase selection
- Idea: Selection method for uncompromising, continous-valued symbolic regression problems[3]
- Increases genetic diversity inside the population[4]
- Higher chance for "specialist" solutions to be selected since it is decided on a per case basis
- Performance increases have been demosntrated in many benchmarking problems[5]

[3] Helmuth, Spector and Matheson (2015), p.12
[4] Helmuth, Spector and Matheson (2015), p.1
[5] La Cava, Spector and Danai (2016), p.744-745

Symbolic Regression

- Task: Find a mathematical model that fits a given set of datapoints
- One of the first applications of Genetic Programming introduced

...

Generalization

Motivation

- Generalization

Genetic Programming

. . .

. . .

Experimental Study

Experimental Study

. . .

Research Design

. . .

Genetic Programming Configuration

...

Results

Descriptive Statistics

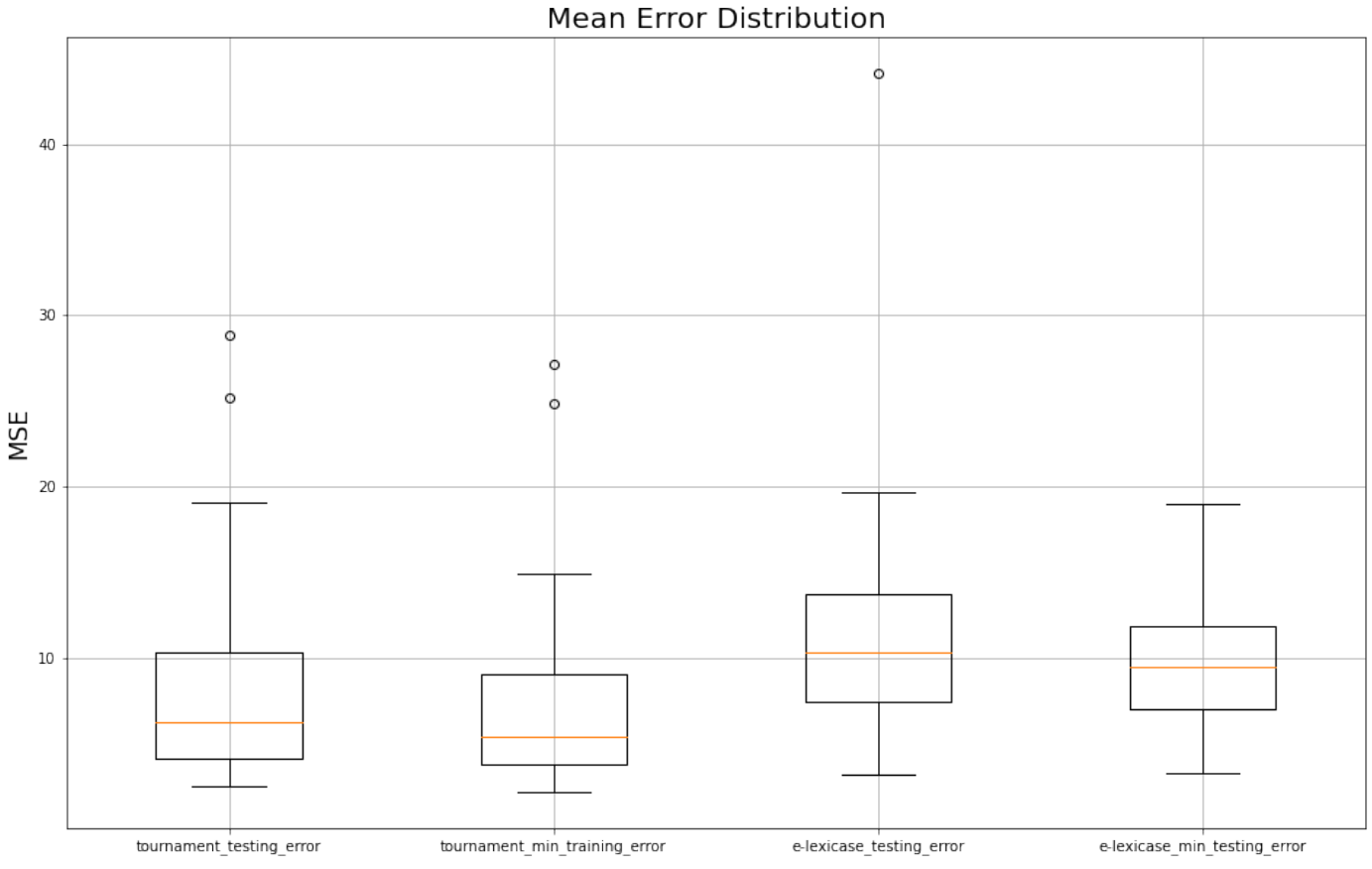

Figure 1: Distribution of Errors

...

Conclusions

Conclusions

. . .

Limitations and open Questions

Limitations and open Questions

...

Fang, Y. and Li, J. (2010) 'A review of tournament selection in genetic programming', in Cai, Z. et al. (eds) *Advances in computation and intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 181–192.

Helmuth, T., Spector, L. and Matheson, J. (2015) 'Solving uncompromising problems with lexicase selection', *IEEE Transactions on Evolutionary Computation*, 19(5), pp. 630–643. doi:10.1109/TEVC.2014.2362729.

Koza, J.R. (1992) *Genetic programming: On the programming of computers by means of natural selection*. Cambridge, MA, USA: MIT Press. Available at: http://mitpress.mit.edu/books/genetic-programming.

La Cava, W., Spector, L. and Danai, K. (2016) 'Epsilon-lexicase selection for regression', in *Proceedings of the genetic and evolutionary computation conference 2016*. New York, NY, USA: