Complementarity-Guided Epsilon Lexicase Selection for Genetic Programming in Symbolic Regression

Hengzhe Zhang, Qi Chen, Bing Xue, Wolfgang Banzhaf, Mengjie Zhang

May 30, 2025

Outline

Introduction

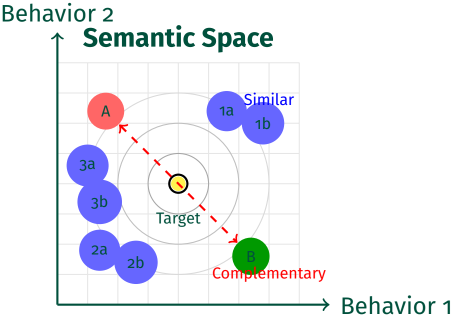
The Problem with Independent Selection

Current Approach:

- Selection operators (tournament, lexicase) select parents independently
- Crossover applied with high probability (80-90%)
- No consideration of parent interaction during crossover

Consequences:

- Parents may be semantically similar or identical
- Crossover fails to combine diverse strengths
- Computational resources wasted
- Evolution can stagnate



Complementarity is important!

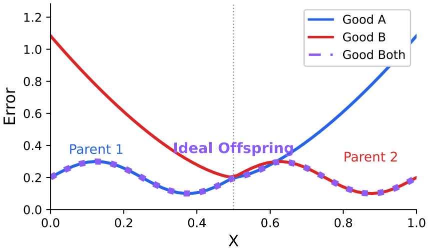
Key Insight: Selection should be aware of crossover behavior!

Ideal Scenario:

- Parent 1 excels on cases A, struggles on cases B
- Parent 2 excels on cases B, struggles on cases A
- Crossover combines strengths from both
- Offspring potentially excels on A **and** B

Challenge:

- Tournament selection favors generalists
- Hard to find truly complementary pairs
- Need specialists that excel in different areas



Background

Tournament Selection:

- Random subsets, select fittest
- Favors generalists

Lexicase Selection [1]:

- Random test case ordering
- Filter by case-specific thresholds
- Favors specialists

Complementary Phenotype Selection [2]:

- First parent: roulette wheel/tournament selection
- Second parent: maximize complementarity
- Limited by generalist first parent

Our Contribution:

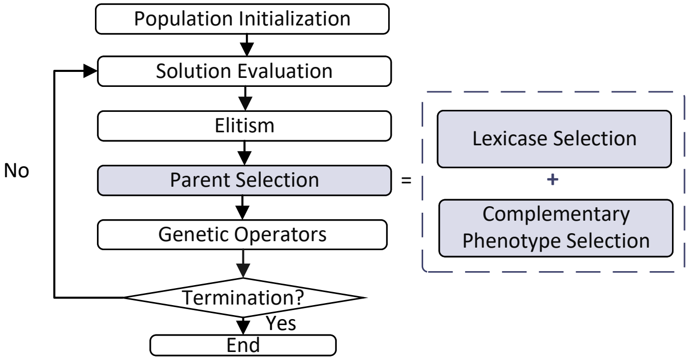
- First parent: lexicase selection
- Second parent: complementarity-guided
- Specialists paired with complementary specialists

[1] La Cava, William et al., 2019“, ‘A probabilistic and multi-objective analysis of lexicase selection and ε-lexicase selection’”, *Evolutionary Computation*

[2] Dolin, Brad, Maribel García Arenas, and Juan J Merelo, 2002“, ‘Opposites attract: Complementary phenotype selection for crossover in genetic programming’”, *Parallel Problem Solving from Nature—PPSN VII*

Proposed Method

Algorithm Framework



Key Features:

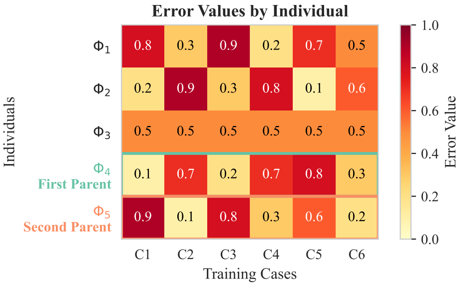
- Standard GP framework with novel selection strategy
- Linear scaling with regularization for fitness evaluation
- Focus on parent complementarity during selection

Step 1 - First Parent Selection:

- Use automatic epsilon-lexicase selection
- Adaptive threshold:
  $\tau_t = \min(e_t) + \text{median}(|e_t - \text{median}(e_t)|)$
- Selects specialists, not generalists

Step 2 - Second Parent Selection:

- Evaluate complementarity with first parent
- Minimize:
  $\mathcal{L}(\Phi_a, \Phi_b) = \frac{1}{m}\sum_{j=1}^{m} \min(E_a[j], E_b[j])$
- Choose parent that best complements Φ_a



Parent selection based on case-wise error patterns

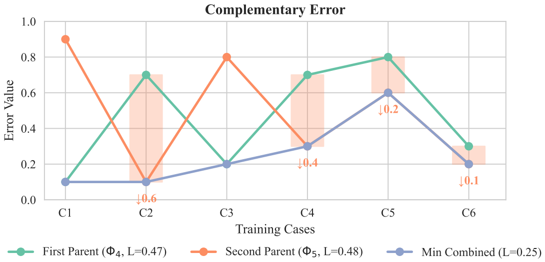
Complementarity Metric:

$$\mathcal{L}(\Phi_a, \Phi_b) = \frac{1}{m} \sum_{j=1}^{m} \min(E_a[j], E_b[j])$$

Interpretation:

- $E_a[j]$, $E_b[j]$: errors on case j
- $\min(E_a[j], E_b[j])$: best performance on case j
- $\mathcal{L}$: average of best case-wise performances
- Lower $\mathcal{L}$ = better complementarity

Tie-breaking: Prefer lower total error



Ideal combination leverages strengths of both parents

Experimental Setup

Datasets:

- 98 regression datasets from PMLB (Fewer than 2000 instances)
- 80:20 train/test split
- Features standardized

GP Parameters:

- Population: 200
- Generations: 100
- Max depth: 10
- Crossover/Mutation: 0.9/0.1
- Functions: +, -, ×, AQ, $\sqrt{\cdot}$, etc.

Baseline Methods:

- **Tournament-3/7**: Tournament selection
- **Lexicase**: Automatic epsilon-lexicase
- **CPS**: Complementary phenotype selection

Evaluation:

- R^2 coefficient of determination
- 30 independent runs per dataset
- Wilcoxon signed-rank test
- Significance level: 0.05

Experimental Results

Table: Statistical comparison of **training** R^2 **scores**

	Lexicase	**Tournament-3**	**Tournament-7**	**CPS**
CLS	64(+)/34(∼)/0(-)	78(+)/18(∼)/2(-)	66(+)/25(∼)/7(-)	36(+)/61(∼)/1(-)
Lexicase	—	67(+)/18(∼)/13(-)	44(+)/36(∼)/18(-)	8(+)/69(∼)/21(-)
Tournament-3	—	—	0(+)/78(∼)/20(-)	1(+)/36(∼)/61(-)
Tournament-7	—	—	—	10(+)/63(∼)/25(-)

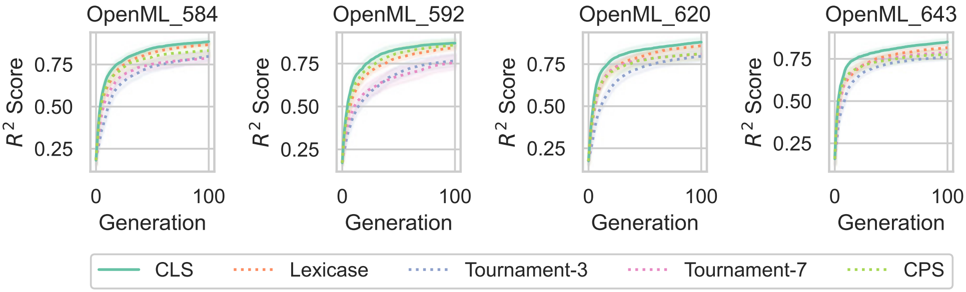
- **Significant improvements**: CLS outperforms baselines on majority of datasets
- **vs Tournament**: 78/98 and 66/98 datasets improved
- **vs Lexicase**: 64/98 datasets improved, 0 datasets worse
- **vs CPS**: 36/98 datasets improved, consistent performance

Table: Statistical comparison of **test** R^2 **scores**

	Lexicase	Tournament-3	Tournament-7	CPS
CLS	23(+)/74(∼)/1(-)	56(+)/39(∼)/3(-)	57(+)/40(∼)/1(-)	25(+)/72(∼)/1(-)
Lexicase	—	55(+)/41(∼)/2(-)	49(+)/48(∼)/1(-)	8(+)/89(∼)/1(-)
Tournament-3	—	—	0(+)/94(∼)/4(-)	0(+)/65(∼)/33(-)
Tournament-7	—	—	—	1(+)/68(∼)/29(-)

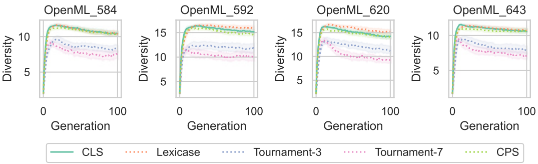
- **Generalization improvement**: Strong test performance across methods
- **Robust performance**: Very few cases where CLS performs worse
- **Consistent gains**: Improvements maintained on unseen data

Evolutionary Performance Trajectories

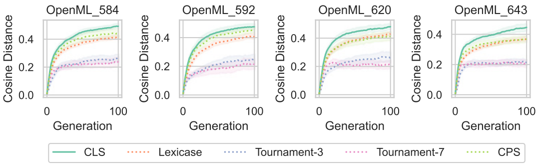


- **Consistent advantage:** CLS performs better from early generations
- **Sustained improvement:** Advantage maintained throughout evolution

Population Diversity Analysis

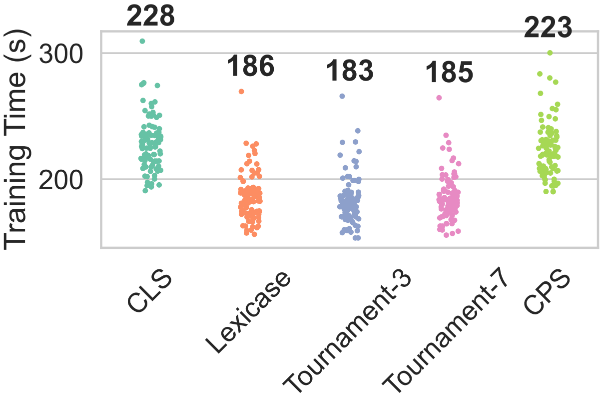


Euclidean Distance Diversity



Cosine Distance Diversity

- **High diversity maintained:** CLS preserves population diversity
- **Functional diversity:** Higher cosine diversity indicates distinct functional behaviors



Time Complexity:

- CLS: $\mathcal{O}(|P|^2 N)$
- Lexicase: $\mathcal{O}(|P|N)$
- Tournament: $\mathcal{O}(|P|)$

Practical Impact:

- Evaluation dominates runtime
- Selection overhead is relatively small

Key Finding: Despite higher theoretical complexity, the practical overhead is relatively small.

Conclusions

Key Contributions and Findings

- **Problem Identified**: Independent parent selection ignores crossover dynamics
- **Solution Proposed**: Complementarity-Guided Lexicase Selection (CLS)
  - First parent: lexicase selection (specialist)
  - Second parent: complementarity-guided selection
  - Leverages case-wise performance differences
- **Experimental Results**:
  - Significant improvements across 98 PMLB datasets
  - Superior training and generalization performance
  - Higher functional diversity in evolved populations
  - Relatively small computational overhead in practice

Future Research Directions

- **Lexicase Selection $\rightarrow$ Batch Lexicase Selection** [1]
- **Lexicase Selection $\rightarrow$ DALex** [2]
- **Lexicase Selection $\rightarrow$ PLex** [3]

[1] Aenugu, Sneha and Lee Spector, 2019", 'Lexicase selection in learning classifier systems'", *Proceedings of the Genetic and Evolutionary Computation Conference*

[2] Ni, Andrew, Li Ding, and Lee Spector, 2024", 'Dalex: Lexicase-like selection via diverse aggregation'", *European Conference on Genetic Programming (Part of EvoStar)*

[3] Ding, Li, Edward Pantridge, and Lee Spector, 2023", 'Probabilistic lexicase selection'", *Proceedings of the Genetic and Evolutionary Computation Conference*

Thank You!

Questions & Discussion

Hengzhe Zhang, Qi Chen, Bing Xue,
Wolfgang Banzhaf, Mengjie Zhang

hengzhe.zhang@ecs.vuw.ac.nz