

Operon C++: an Efficient Genetic Programming Framework for Symbolic Regression

Supplementary Material

Bogdan Burlacu

Gabriel Kronberger

Michael Kommenda

1 Statistical tests

Statistical tests are performed for modeling error (training and test NMSE) and elapsed time using the 50 repetitions of each experiment configuration. We use the Mann-Whitney-U test (two-sided, $\alpha = 0.05$) to ascertain whether two distributions are different and then compare median values to determine the direction of the relationship. Significant differences ($p < \alpha$) are marked in bold font and the Tables 1, 2, 3 are read row-wise.

1.1 Modeling error

We use the normalized mean squared error (NMSE) defined as:

$$\text{NMSE}(\hat{y}, y) = \frac{\text{MSE}(\hat{y}, y)}{\text{Var}(y)}$$

where \hat{y} is the model prediction, y is the actual target value, MSE is the mean squared error and Var is the variance.

1.2 Elapsed time

When testing DEAP and HeuristicLab, we observed that the best runtime performance was obtained by using a coarse-grained parallel model. That is, instead of parallelizing fitness evaluation inside the evolutionary process, we parallelize the entire experiment by performing the 50 repetitions in parallel. This means that we obtained two types of runtime measurements:

1. The total experiment time T elapsed for the $N = 50$ repetitions run in parallel
2. Elapsed times t_i for each individual (single-threaded) run

For DEAP and HeuristicLab, we then report in the paper the average time calculated as $\bar{T} = \frac{T}{N}$. An alternative way to report execution times (since a reviewer also requested IQR or Stddev of the time) would be to calculate an *adjusted runtime* as follows:

- For each configuration, we calculate

$$\bar{t} = \frac{1}{N} \sum_{i=1}^N t_i$$

- Then we calculate adjusted run execution times

$$t_i^* = \bar{T} \cdot \frac{t_i}{\bar{t}}$$

The advantage of using adjusted runtimes is that we preserve the distribution, which means that we can report median \pm IQR and perform statistical tests. For *Operon*, the concurrency model allows fine-grained parallelism as the offspring population is generated concurrently, so no adjustment is necessary.

2 Summary of the results

We use the term “outperforms” to refer to a statistically significant difference ($p < 0.05$) combined with the condition that the first median is lower than the second median and the absolute difference is greater than 0.001. Bear in mind that 50 repetitions is not a very large sample size.

2.1 Training performance

DEAP

- Outperforms *HeuristicLab* on 1/9 problems (Poly-10).

HeuristicLab

- Outperforms *Operon* on 1/9 problems (Airflow Self-Noise).
- Outperforms *DEAP* on 6/9 problems.

Operon

- Outperforms *DEAP* on all problems.
- Outperforms *HeuristicLab* on 5/9 problems.

2.2 Test performance

DEAP

- Does not outperform the other frameworks on any problem.

HeuristicLab

- Outperforms *Operon* on 1/9 problems (Airflow Self-Noise).
- Outperforms *DEAP* on 5/9 problems.

Operon

- Outperforms *DEAP* on 7/9 problems.
- Outperforms *HeuristicLab* on 4/9 problems.

2.3 Elapsed time

In terms of runtime *Operon* is always faster so we don’t include it in the comparison. The single-precision variant is faster than the double-precision variant. *DEAP* is faster than *HeuristicLab* on two problems (Friedman-I and Friedman-II), while *HeuristicLab* is faster than *DEAP* on six problems (all except Friedman-I, Friedman-II and GP-Challenge).

	DEAP	HeuristicLab	Operon (double)	Operon (float)
Airflow Self-Noise				
DEAP	–	–	–	–
HeuristicLab	2.6e-10	–	7.0e-03	–
Operon (double)	1.3e-06	–	–	–
Operon (float)	1.8e-08	–	–	–
Breiman-I				
DEAP	–	–	–	–
HeuristicLab	–	–	–	–
Operon (double)	2.9e-03	1.8e-03	–	–
Operon (float)	1.5e-03	1.3e-03	–	–
Chemical-I				
DEAP	–	–	–	–
HeuristicLab	6.1e-06	–	–	–
Operon (double)	7.6e-10	8.1e-04	–	–
Operon (float)	7.7e-10	2.5e-03	–	–
Concrete Compressive Strength				
DEAP	–	–	–	–
HeuristicLab	1.1e-04	–	–	–
Operon (double)	1.6e-03	–	–	–
Operon (float)	3.8e-06	–	–	–
Friedman-I				
DEAP	–	–	–	–
HeuristicLab	9.9e-16	–	–	–
Operon (double)	6.0e-15	–	–	–
Operon (float)	9.6e-13	–	–	–
Friedman-II				
DEAP	–	–	–	–
HeuristicLab	5.0e-11	–	–	–
Operon (double)	1.3e-12	–	–	–
Operon (float)	1.1e-10	–	–	–
GP-Challenge				
DEAP	–	–	–	–
HeuristicLab	1.0e-10	–	–	–
Operon (double)	1.8e-14	2.7e-04	–	–
Operon (float)	2.2e-13	6.6e-03	–	–
Poly-10				
DEAP	–	4.6e-03	–	–
HeuristicLab	–	–	–	–
Operon (double)	3.6e-02	3.7e-06	–	–
Operon (float)	2.7e-02	7.7e-07	–	–
Spatial Coevolution				
DEAP	–	–	–	–
HeuristicLab	–	–	–	–
Operon (double)	4.6e-05	1.2e-04	–	–
Operon (float)	3.5e-04	1.6e-03	–	–

Table 1: Training performance p-values

	DEAP	HeuristicLab	Operon (double)	Operon (float)
Airflow Self-Noise				
DEAP	–	–	–	–
HeuristicLab	3.7e-09	–	4.9e-02	–
Operon (double)	1.4e-05	–	–	–
Operon (float)	3.6e-07	–	–	–
Breiman-I				
DEAP	–	–	–	–
HeuristicLab	–	–	–	–
Operon (double)	1.6e-03	2.9e-03	–	–
Operon (float)	8.4e-04	5.4e-03	–	–
Chemical-I				
DEAP	–	–	–	–
HeuristicLab	–	–	–	–
Operon (double)	–	–	–	–
Operon (float)	–	–	–	–
Concrete Compressive Strength				
DEAP	–	–	–	–
HeuristicLab	–	–	–	–
Operon (double)	–	–	–	–
Operon (float)	–	–	–	–
Friedman-I				
DEAP	–	–	–	–
HeuristicLab	9.6e-15	–	–	–
Operon (double)	2.8e-14	–	–	–
Operon (float)	1.1e-11	–	–	–
Friedman-II				
DEAP	–	–	–	–
HeuristicLab	4.3e-11	–	–	–
Operon (double)	1.1e-12	–	–	–
Operon (float)	1.1e-10	–	–	–
GP-Challenge				
DEAP	–	–	–	–
HeuristicLab	7.1e-09	–	–	–
Operon (double)	2.4e-14	8.0e-05	–	–
Operon (float)	3.5e-13	4.9e-03	–	–
Poly-10				
DEAP	–	–	–	–
HeuristicLab	–	–	–	–
Operon (double)	1.4e-02	3.1e-05	–	–
Operon (float)	1.1e-02	1.8e-05	–	–
Spatial Coevolution				
DEAP	–	–	–	–
HeuristicLab	2.0e-05	–	–	–
Operon (double)	3.9e-08	1.7e-02	–	–
Operon (float)	1.1e-11	2.0e-04	–	–

Table 2: Test performance p-values

	DEAP	HeuristicLab	Operon (double)	Operon (float)
Airflow Self-Noise				
DEAP	–	–	–	–
HeuristicLab	4.5e-18	–	–	–
Operon (double)	3.5e-18	3.5e-18	–	–
Operon (float)	3.5e-18	3.5e-18	4.5e-12	–
Breiman-I				
DEAP	–	–	–	–
HeuristicLab	4.7e-02	–	–	–
Operon (double)	3.5e-18	3.5e-18	–	–
Operon (float)	3.5e-18	3.5e-18	1.7e-11	–
Chemical-I				
DEAP	–	–	–	–
HeuristicLab	3.5e-18	–	–	–
Operon (double)	3.5e-18	3.5e-18	–	–
Operon (float)	3.5e-18	3.5e-18	1.5e-07	–
Concrete Compressive Strength				
DEAP	–	–	–	–
HeuristicLab	3.5e-18	–	–	–
Operon (double)	3.5e-18	3.5e-18	–	–
Operon (float)	3.5e-18	3.5e-18	4.0e-10	–
Friedman-I				
DEAP	–	4.4e-03	–	–
HeuristicLab	–	–	–	–
Operon (double)	9.7e-14	4.5e-17	–	–
Operon (float)	3.5e-18	3.5e-18	1.7e-11	–
Friedman-II				
DEAP	–	4.9e-02	–	–
HeuristicLab	–	–	–	–
Operon (double)	5.7e-18	3.8e-18	–	–
Operon (float)	3.5e-18	3.5e-18	3.5e-16	–
GP-Challenge				
DEAP	–	–	–	–
HeuristicLab	–	–	–	–
Operon (double)	2.1e-12	3.3e-16	–	–
Operon (float)	4.8e-18	3.5e-18	2.8e-12	–
Poly-10				
DEAP	–	–	–	–
HeuristicLab	3.5e-18	–	–	–
Operon (double)	3.5e-18	3.5e-18	–	–
Operon (float)	3.5e-18	3.5e-18	9.8e-10	–
Spatial Coevolution				
DEAP	–	–	–	–
HeuristicLab	5.0e-17	–	–	–
Operon (double)	3.5e-18	3.5e-18	–	–
Operon (float)	3.5e-18	3.5e-18	1.4e-14	–

Table 3: Elapsed time p-values