Fitness Predictors in Genetic Programming

Jan Mayer

Problem

- Solve symbolic regression using GP
- In each generation all solutions have to be evaluated in all points of the dataset
- Is there a way to reduce this effort?

Fitness Predictors

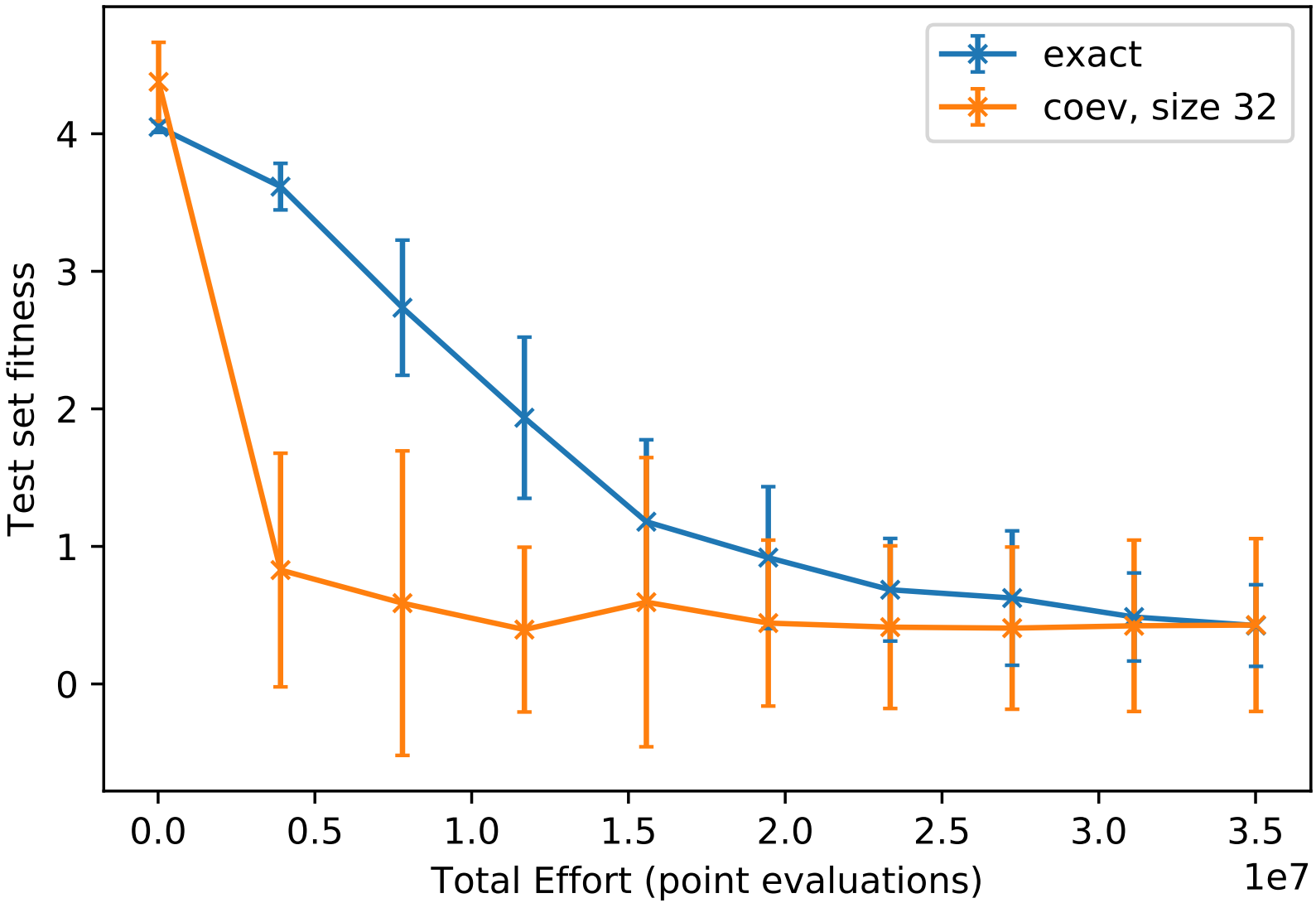
- A (small) subset of the dataset
- Instead of evaluating in all points, only use the points in the current fitness predictor
- But how to choose the predictors?

Coevolution

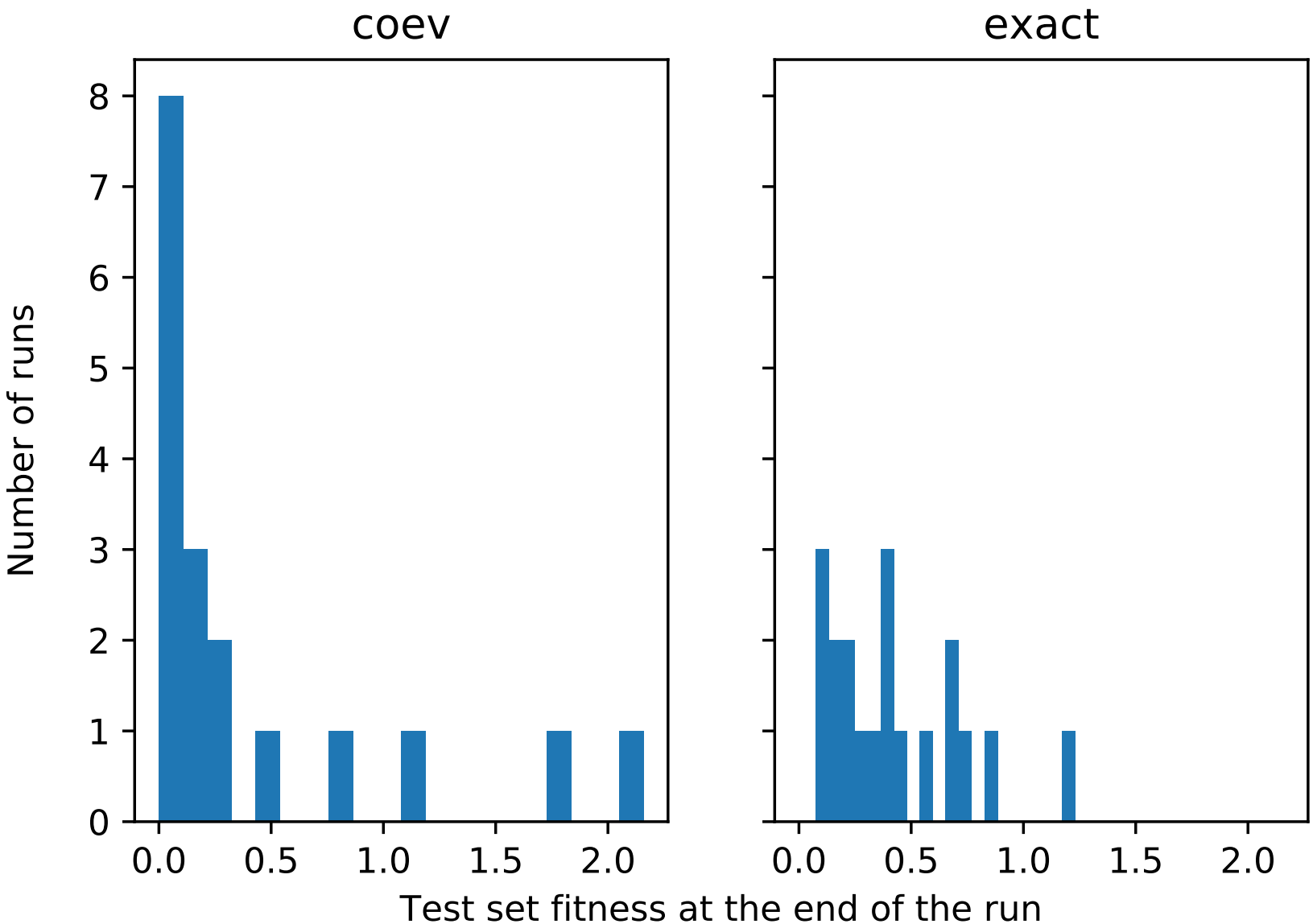
Michael D. Schmidt & Hod Lipson published a paper, in which they introduce fitness predictors constructed using coevolution. The main idea is:

- Aside from the main population have another one with predictors
- Each generation, evaluate solutions using top-ranked predictor
- Predictors that predict the fitness of the solutions more accurately are ranked higher
- Predictors are evolved using standard crossover, mutation and selection
- This way more and more accurate predictors are used as the evolution runs

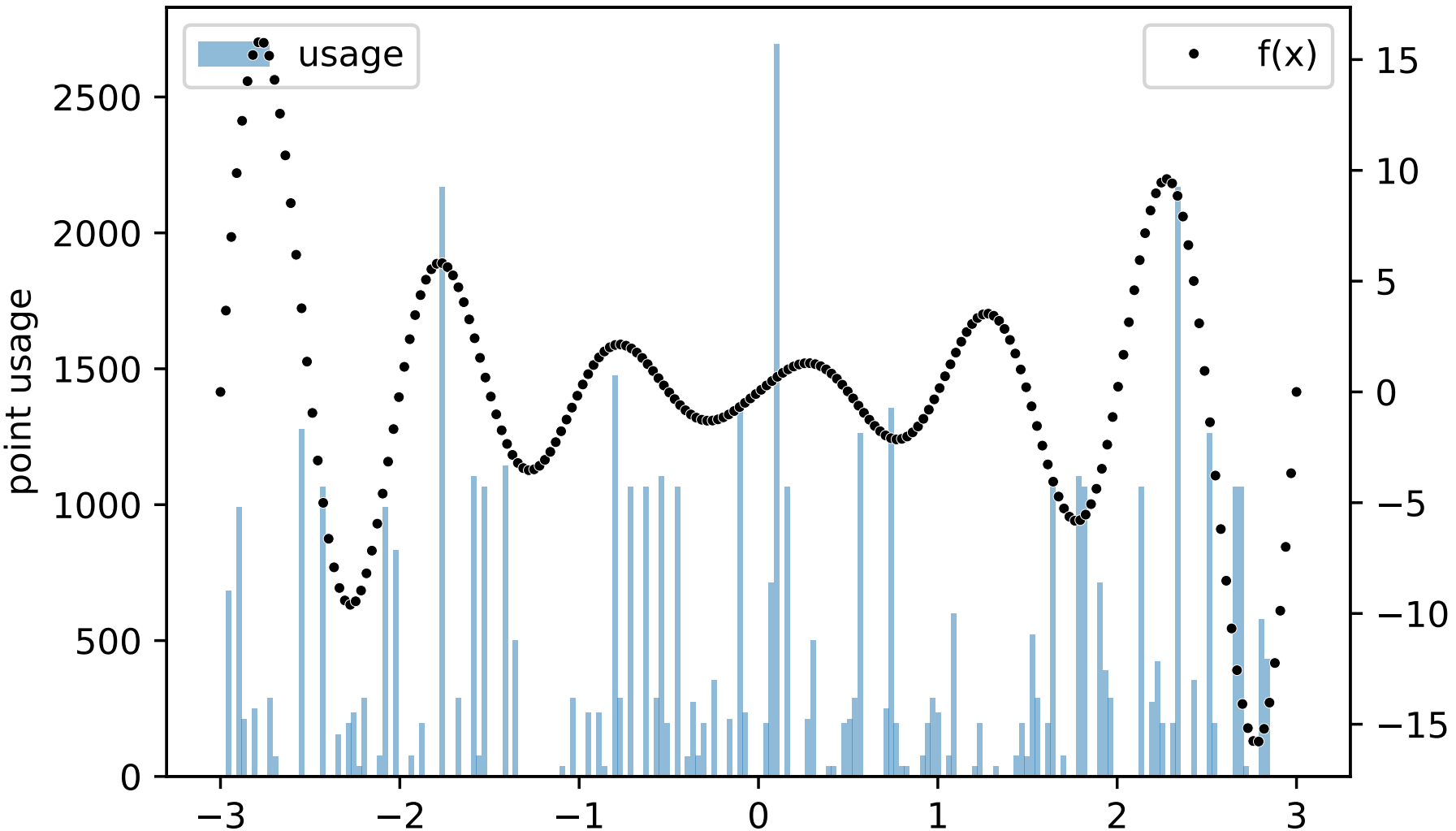
Results


exact
coev, size 32
Test set fitness
0
1
2
3
4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Total Effort (point evaluations)
1e7

Results


coev
exact
Number of runs
0
1
2
3
4
5
6
7
8
0.0
0.5
1.0
1.5
2.0
Test set fitness at the end of the run

Results


usage
f(x)
point usage
2500
2000
1500
1000
500
0
15
10
5
0
−5
−10
−15
−3
−2
−1
0
1
2
3

Future work

- Design our own methods of constructing the predictors. Is coevolution unnecessarily complicated for this task?
- Compare them with the known methods