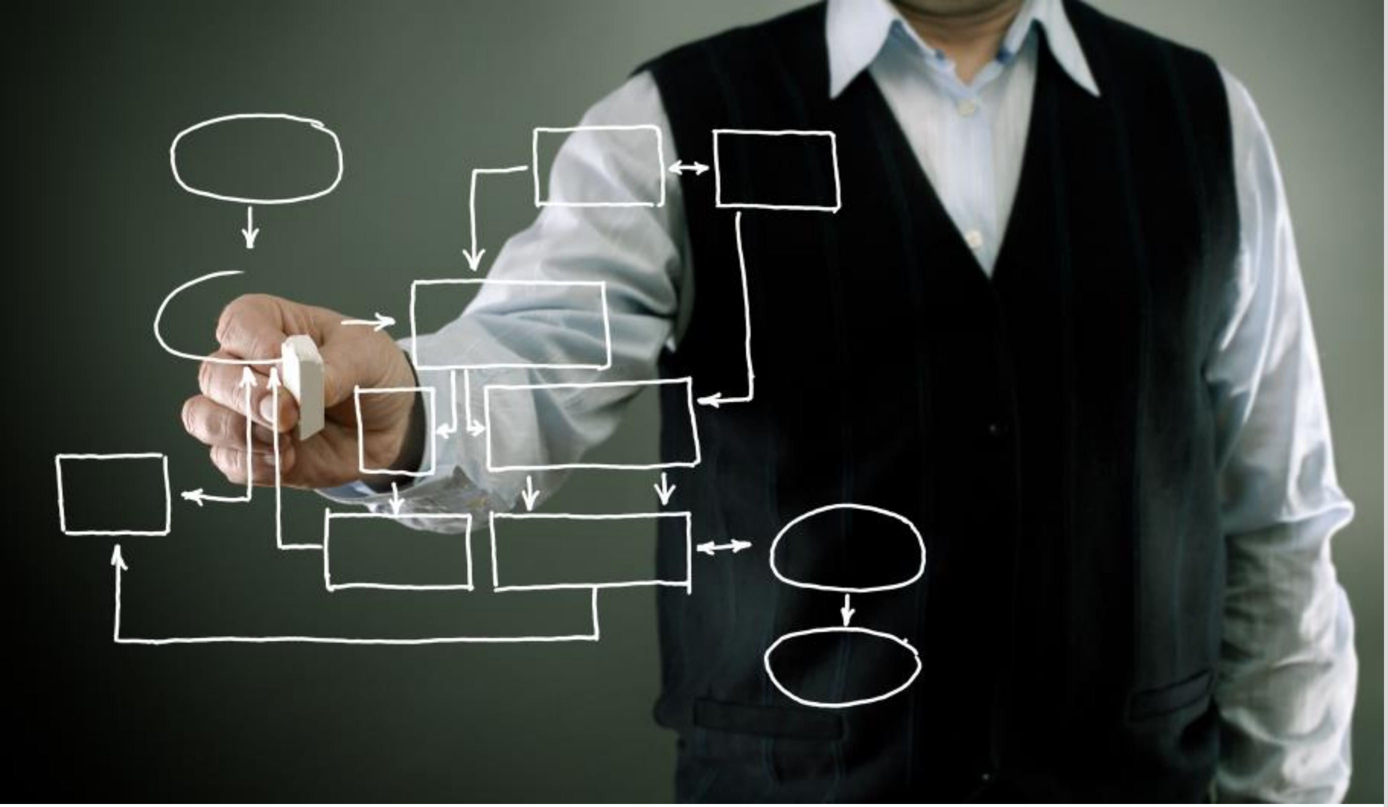


Effects of Constant Optimization by Nonlinear Least Squares Minimization in Symbolic Regression

Michael Kommenda, **Gabriel Kronberger**, Stephan Winkler, Michael Affenzeller, and Stefan Wagner

HEAL
HEURISTIC AND EVOLUTIONARY ALGORITHMS LABORATORY

GECCO
HeuristicLab
Josef Ressel-Zentrum HEUREKA!

Contact:
Michael Kommenda
Heuristic and Evolutionary Algorithms Lab (HEAL)
Softwarepark 11
A-4232 Hagenberg

e-mail:
michael.kommenda@fh-hagenberg.at
Web:
http://heal.heuristiclab.com
http://heureka.heuristiclab.com

Symbolic Regression

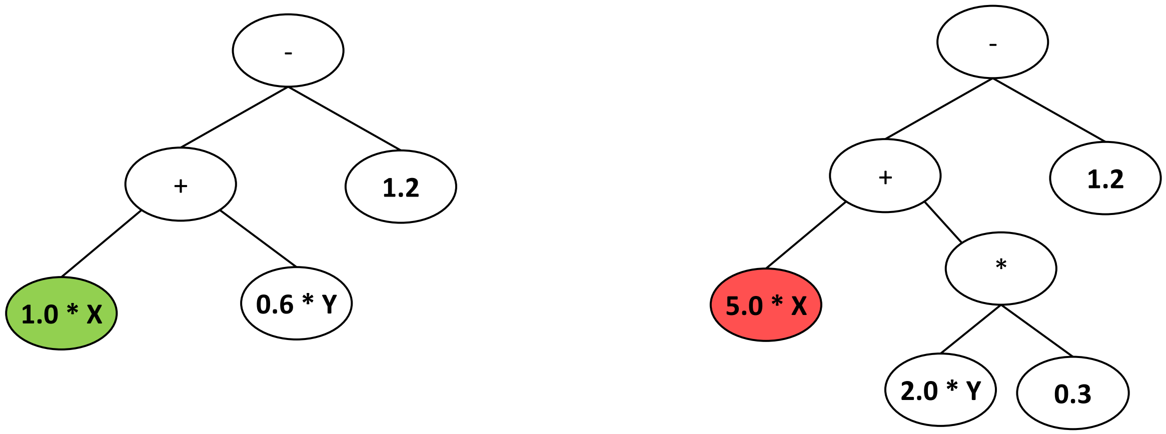
- **Model a relationship between input variables x and target variable y without any predefined structure**

$$y = f(x, w) + \varepsilon$$

- **Minimization of ε using an evolutionary algorithm**
  - Model structure
  - Used variables
  - **Constants / weights**

Research Assumption

The correct model structure is found during the algorithm execution, but not recognized due to misleading / wrong constants.



Constants in Symbolic Regression

Ephemeral Random Constants

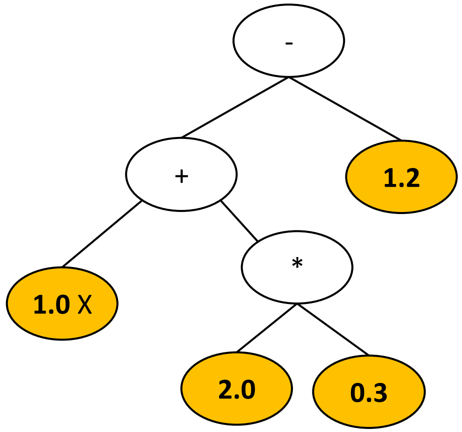
- Randomly initialized constants
- Remain fixed during the algorithm run

Evolutionary Constants

- Updated by mutation
  - $C_{new} = C_{old} + N(0, \sigma)$
  - $C_{new} = C_{old} * N(1, \sigma)$

Finding correct constants

- combination of existing values
- mutation of constant symbol nodes
  - undirected changes to values



Summary of Previous Research

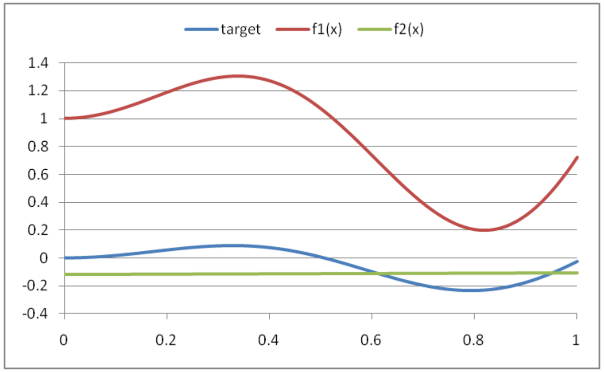
- Faster genetic programming based on local gradient search of numeric leaf values (Topchy and Punch, GECCO 2001)
- Improving gene expression programming performance by using differential evolution (Zhang et al., ICMLA 2007)
- Evolution Strategies for Constants Optimization in Genetic Programming (Alonso, ICTAI 2009)
- Differential Evolution of Constants in Genetic Programming Improves Efficacy and Bloat (Mukherjee and Eppstein, GECCO 2012)

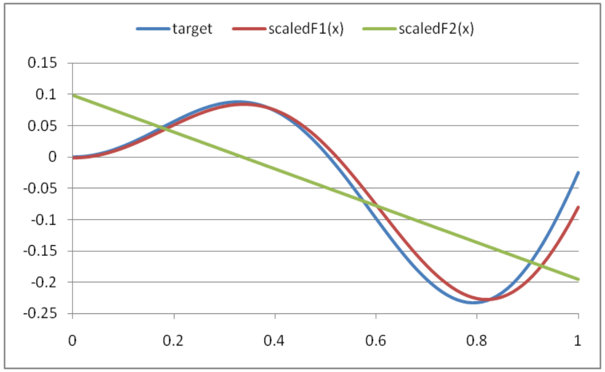
Linear Scaling

- **Improving Symbolic Regression with Interval Arithmetic and Linear Scaling (Keijzer, EuroGP 2003)**

- **Use Pearson's R^2 as fitness function and perform linear scaling**
  - Removes necessity to find correct offset and scale
  - Computationally efficient

- **Outperforms the local gradient search**





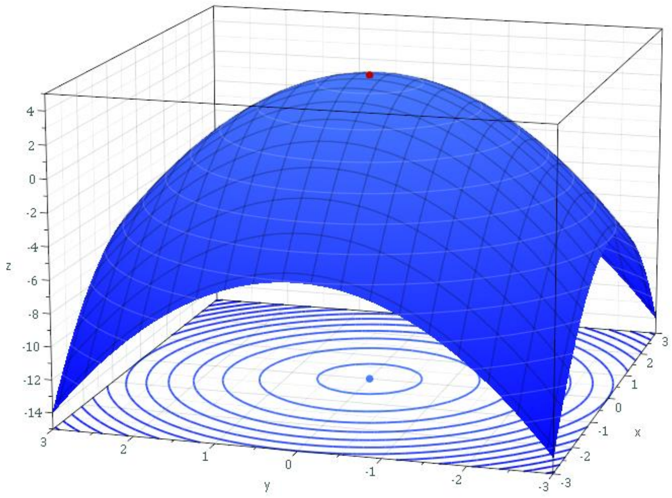
Constant Optimization

Concept

- Treat all constants as parameters
- Local optimization step
- Multidimensional optimization

Levenberg-Marquardt Algorithm

- Least squares fitting of model parameters to empirical data
- $Minimize\ Q(\beta) = \sum_{i=1}^{m}[y_i - f(x_i, \beta)]^2$
- Uses gradient and Jacobian matrix information
- Implemented e.g. by ALGLIB



Gradient Calculation

Transformation of symbolic expression tree

- Extract initial numerical values (starting point)
- Add scaling tree nodes

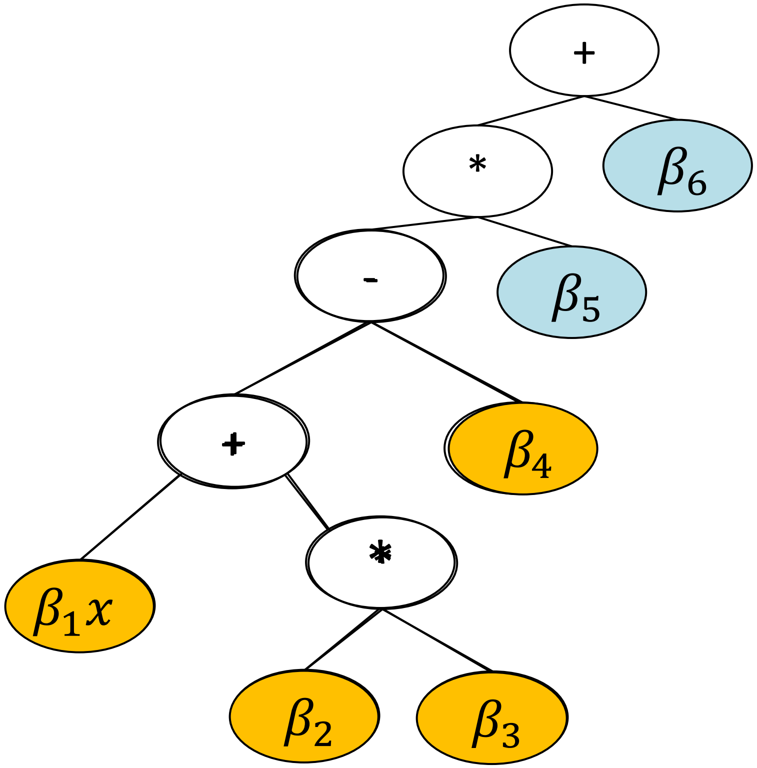
Automatic differentiation

- Provided e.g. by AutoDiff
- Numerical gradient calculation in one pass
- Faster compared to symbolic differentiation

$$\nabla f = \left(\frac{\partial f}{\partial \beta_1}, \frac{\partial f}{\partial \beta_2}, \ldots, \frac{\partial f}{\partial \beta_n} \right)$$

Update tree with optimized values

- Optionally calculate new fitness



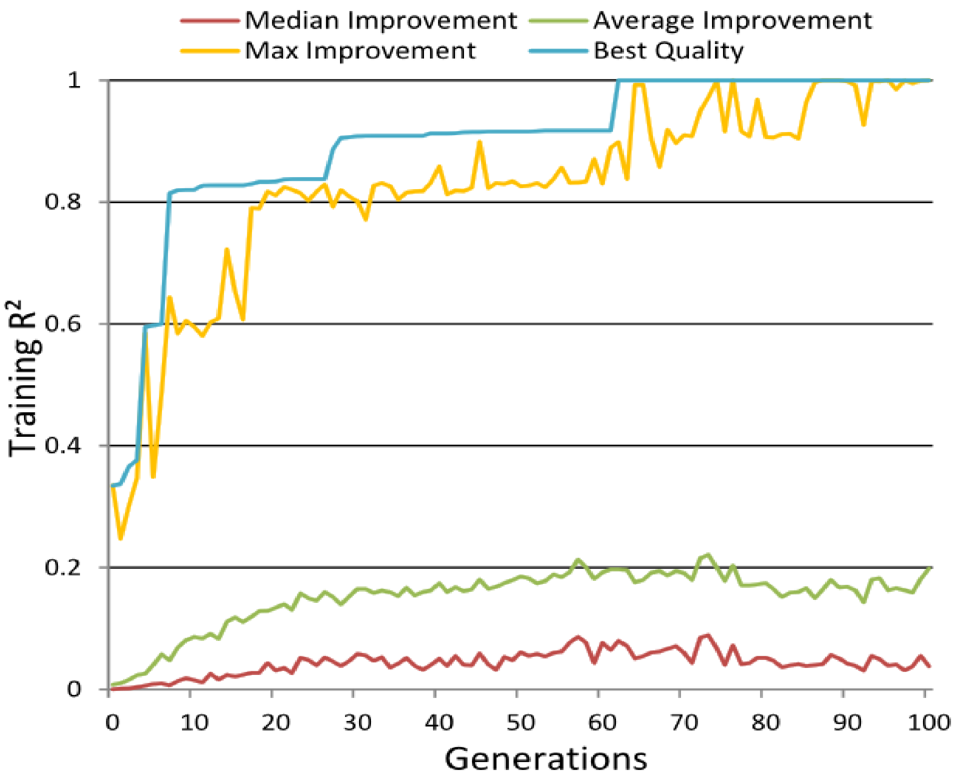
Constant Optimization Improvement

$$Improvement = Quality_{optimized} - Quality_{original}$$

Exemplary GP Run

- Average & median improvement stays constantly low
- Maximum improvement almost reaches the best quality found

- Crossover worsens good individuals
- The quality of few individuals can be dramatically increased



Problems

Symbolic regression benchmarks

- Better GP Benchmarks: Community Survey Results and Proposals (White et al., GPEM 2013)

Problem	Function	Training	Test
Nguyen-7	$f(x) = \ln(x+1) + \ln(x^2+1)$	20	500
Keijzer-6	$f(x,y,z) = \frac{30xz}{(x-10)y^2}$	20	120
Vladislavleva-4	$f(x_1,\dots,x_5) = \frac{10}{5+\sum(x_i-30)^2}$	1024	5000
Pagie-1	$f(x,y) = \frac{1}{1+x^{-4}} + \frac{1}{1+y^{-4}}$	676	1000
Poly-10	$f(x_1,\dots,x_{10}) = x_1x_2 + x_3x_4 + x_5x_6 + x_1x_7x_9 + x_3x_6x_{10}$	250	250
Friedman-2	$f(x_1,\dots,x_{10}) = 10\sin(\pi x_1x_2) + 20(x_3-0.5)^2 + 10x_4 + 5x_5 + N(0,1)$	500	5000
Tower	Real world data	3136	1863

Algorithm Configurations

Genetic Programming with strict offspring selection

- Only child individuals with better quality compared to the fitter parent are accepted in the new generation

Varying parameters

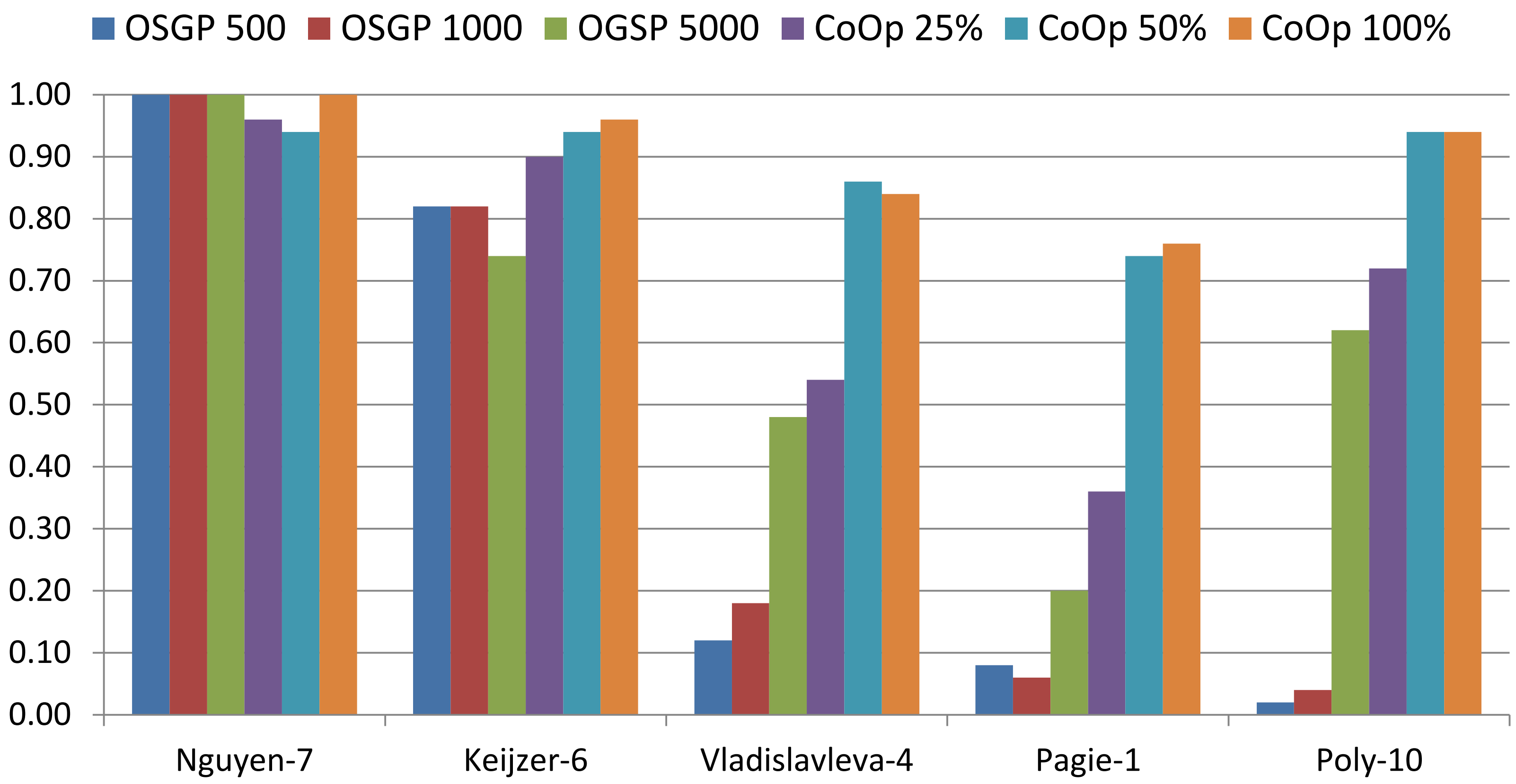
- Population size of 500, 1000, and 5000 for runs without constant optimization
- Probability for constant optimization 25%, 50%, and 100% (population size 500)

All others parameters were not modified

- Maximum selection pressure of 100 was used as termination criterion
- Size constraints of tree length 50 and depth 12
- Mutation rate of 25%
- Function set consists solely of arithmetic functions (except Nguyen-7)

Results - Quality

- Success rate (test $R^2 > 0.99$)



Results - Quality

Noisy datasets

- Success rate not applicable
- R^2 of best training solution ($\mu \pm \sigma$)

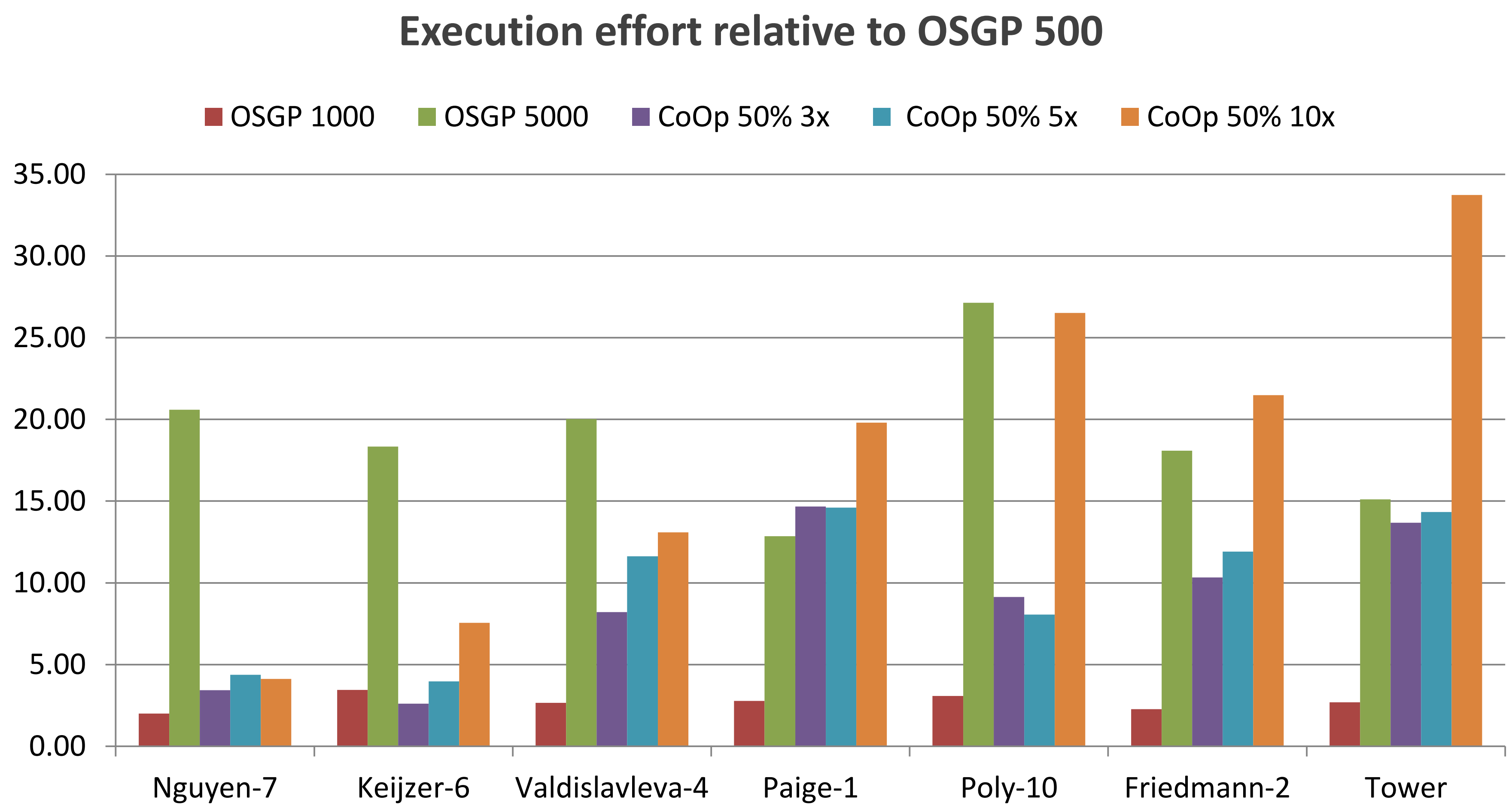
Configuration	Friedman-2		Tower	
	Training	Test	Training	Test
OSGP 500	0.836 ± 0.027	0.768 ± 0.172	0.877 ± 0.007	0.876 ± 0.012
OSGP 1000	0.857 ± 0.036	0.831 ± 0.102	0.880 ± 0.006	0.877 ± 0.024
OSGP 5000	0.908 ± 0.035	0.836 ± 0.191	0.892 ± 0.006	0.890 ± 0.008
CoOp 25%	0.959 ± 0.001	0.871 ± 0.151	0.919 ± 0.006	0.916 ± 0.007
CoOp 50%	0.967 ± 0.000	0.920 ± 0.086	0.925 ± 0.005	0.921 ± 0.006
CoOp 100%	0.964 ± 0.000	0.864 ± 0.142	0.932 ± 0.005	0.927 ± 0.005

Results – LM Iterations

- **Constant optimization probability of 50%**
- **Varying iterations for the LM algorithm (3x, 5x, 10x)**
  - success rate
  - respectively test R^2 for noisy datasets

Problem	OGSP 5000	CoOp 50% 3x	CoOp 50% 5x	CoOp 50% 10x
Nguyen-7	1.00	0.92	0.92	0.94
Keijzer-6	0.74	0.92	0.88	0.94
Vladislavleva-4	0.48	0.56	0.82	0.86
Pagie-1	0.20	0.26	0.52	0.74
Poly-10	0.62	0.78	0.88	0.94
Friedman-2	0.836 ± 0.191	0.946 ± 0.046	0.943 ± 0.076	0.920 ± 0.086
Tower	0.890 ± 0.009	0.902 ± 0.010	0.912 ± 0.008	0.921 ± 0.006

Results - Execution Effort



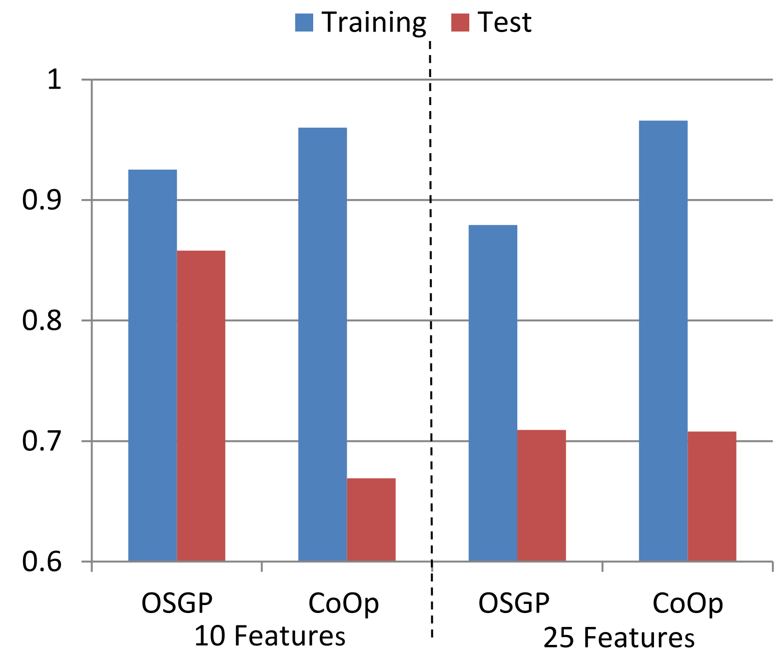
Feature Selection Problems

Artificial datasets

- 100 input variables $N(0,1)$
- Linear combination of 10/25 variables with weights $U(0,10)$
- noisy → max R^2 = 0.90
- Training 120 rows, Test 500 rows
- Population size 500
- Constant optimization 50% 5x

Observation

- Constant optimization can lead to overfitting
- Selection of correct features is also an issue



Conclusion

Constant optimization improves the success rate and quality of models

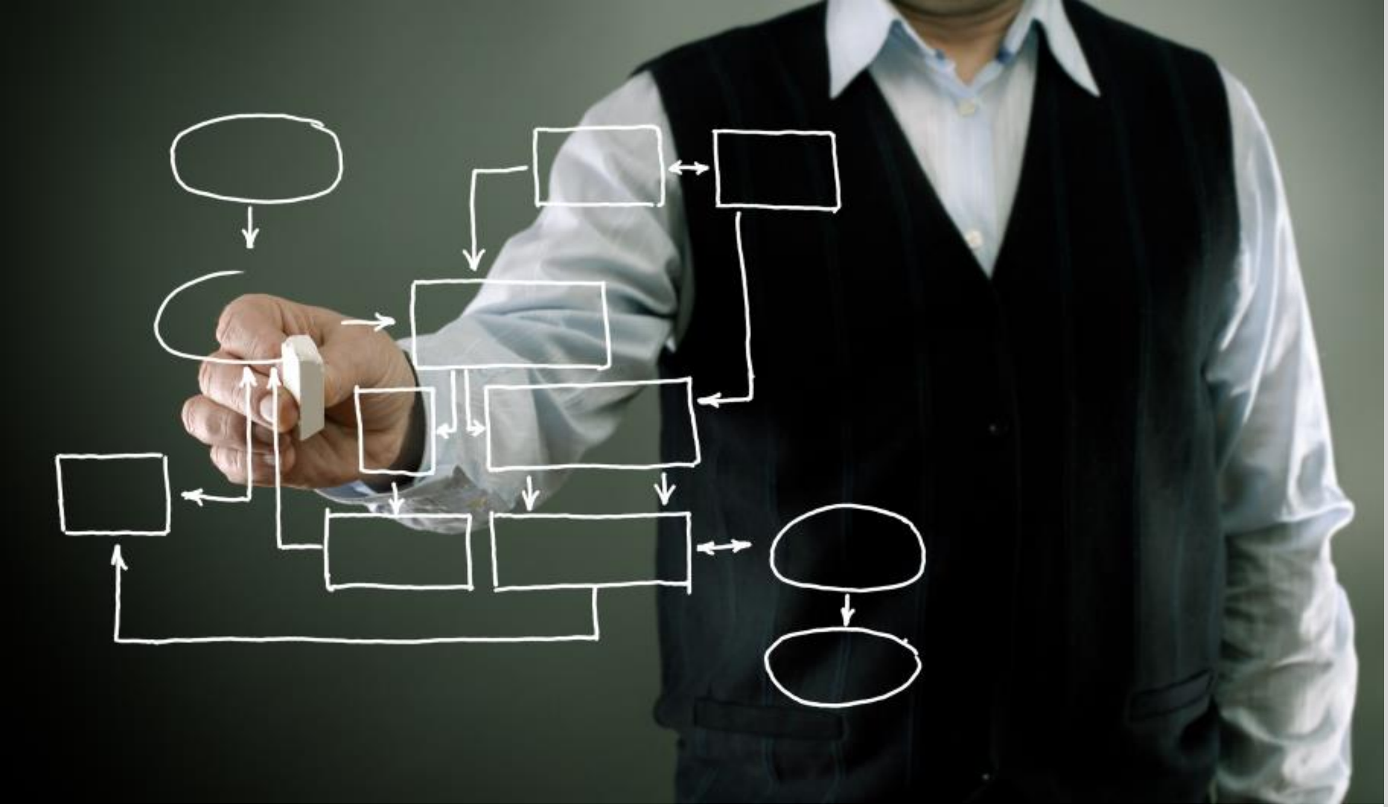
- Better results with smaller population size
- Especially useful for post-processing of models

Removes the effort of evolving correct constants

- Genetic programming can concentrate on the model structure and feature selection

Ready-to-use implementation in HeuristicLab

- Configurable probability, iterations, random sampling
- All experiments available for download
- http://dev.heuristiclab.com/AdditionalMaterial



Effects of Constant Optimization by Nonlinear Least Squares Minimization in Symbolic Regression

Michael Kommenda, **Gabriel Kronberger**, Stephan Winkler, Michael Affenzeller, and Stefan Wagner

HEAL
HEURISTIC AND EVOLUTIONARY
ALGORITHMS LABORATORY

Contact:
Michael Kommenda
Heuristic and Evolutionary
Algorithms Lab (HEAL)
Softwarepark 11
A-4232 Hagenberg

e-mail:
michael.kommenda@fh-hagenberg.at
Web:
http://heal.heuristiclab.com
http://heureka.heuristiclab.com