Empirical Scaling Analyser (ESA)

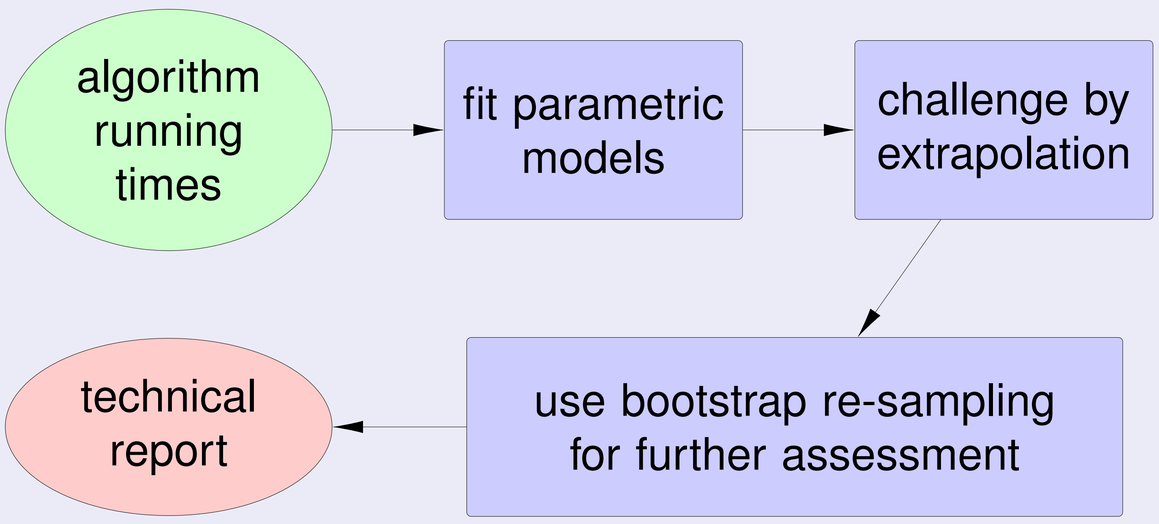
An Automated System for Empirical Analysis of Performance Scaling

Zongxu Mu & Holger H. Hoos · University of British Columbia

What Is ESA?

- Automated tool for performing empirical scaling analysis [1] on algorithm running times
- Methodology has been successfully applied to study SAT [3], TSP [2], etc.
- Available online at `www.cs.ubc.ca/labs/beta/Projects/ESA/esa-online.html`

How Does ESA Work?



Dataset Details

n	200	250	300	350	400	450	500
# Instances	601	589	633	558	579	572	578
mean	0.0065	0.0167	0.0479	0.0743	0.2162	0.2634	2.1713
coefficient of variation	1.9323	2.7076	7.1479	4.6358	8.1654	6.2329	17.9680
Q(0.1)	0.0006	0.0011	0.0016	0.0022	0.0035	0.0050	0.0066
Q(0.25)	0.0010	0.0019	0.0032	0.0043	0.0076	0.0101	0.0144
median	0.0021	0.0045	0.0075	0.0109	0.0182	0.0241	0.0365
Q(0.75)	0.0057	0.0121	0.0210	0.0298	0.0536	0.0867	0.1292
Q(0.9)	0.0157	0.0364	0.0599	0.0891	0.2392	0.3534	0.4375

n	600	700	800	900	1000
# Instances	572	636	584	592	593
mean	2.5027	3.3031	2.7717	15.5353	30.1594
coefficient of variation	13.3185	7.8551	5.1294	6.3333	5.4317
Q(0.1)	0.0124	0.0184	0.0268	0.0359	0.0540
Q(0.25)	0.0240	0.0395	0.0550	0.0801	0.1190
median	0.0564	0.1083	0.1797	0.2668	0.3845
Q(0.75)	0.2014	0.4775	0.7455	1.3348	1.8264
Q(0.9)	1.0791	2.0195	3.3366	8.3035	14.4725

Model Fitting

Fitted models of median running times:

	Model	RMSE (support)	RMSE (challenge)
WalkSAT/SKC	Exp. Model $6.89157 \times 10^{-4} \times 1.00798^{n}$	0.0008564	0.7600
	Poly. Model $8.83962 \times 10^{-11} \times n^{3.18915}$	**0.0007433**	**0.03142**

95% confidence intervals for model parameters:

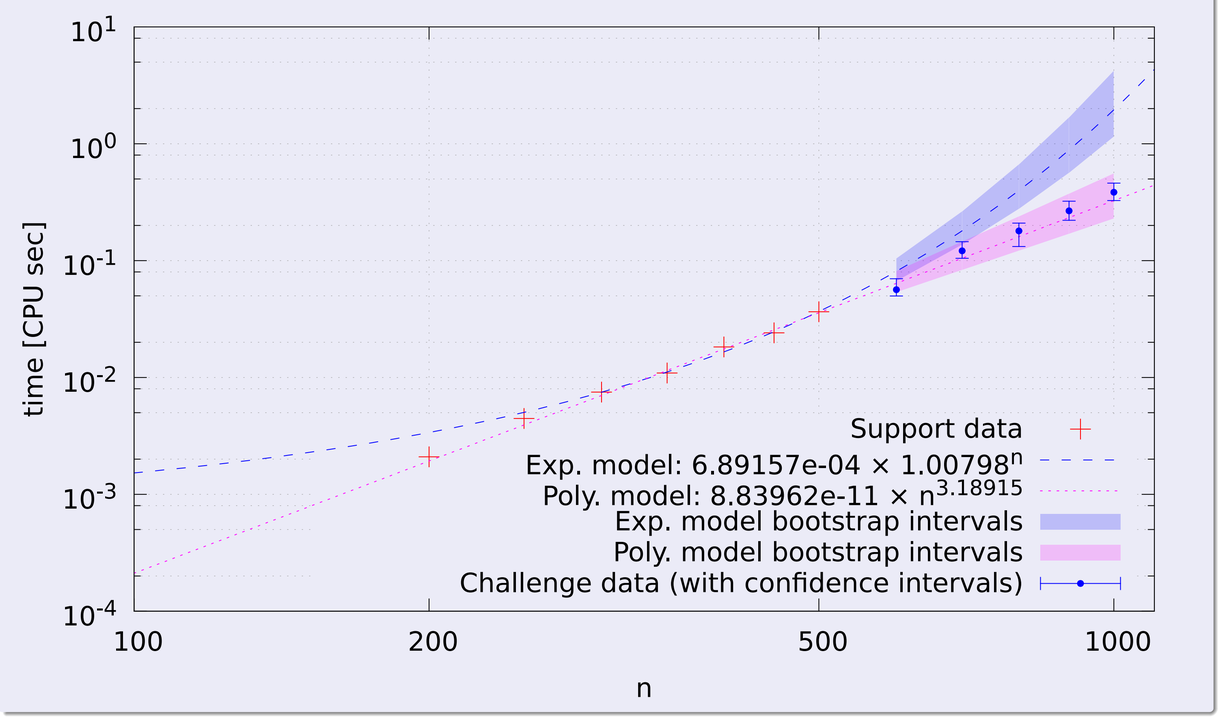
Solver	Model	Conf. interval of *a*	Conf. interval of *b*
WalkSAT/SKC	Poly.	$[2.58600 \times 10^{-12}, 8.63869 \times 10^{-10}]$	$[2.80816, 3.76751]$
	Exp.	$[4.05064 \times 10^{-4}, 1.00662 \times 10^{-3}]$	$[1.00709, 1.00924]$

Challenging Fitted Models

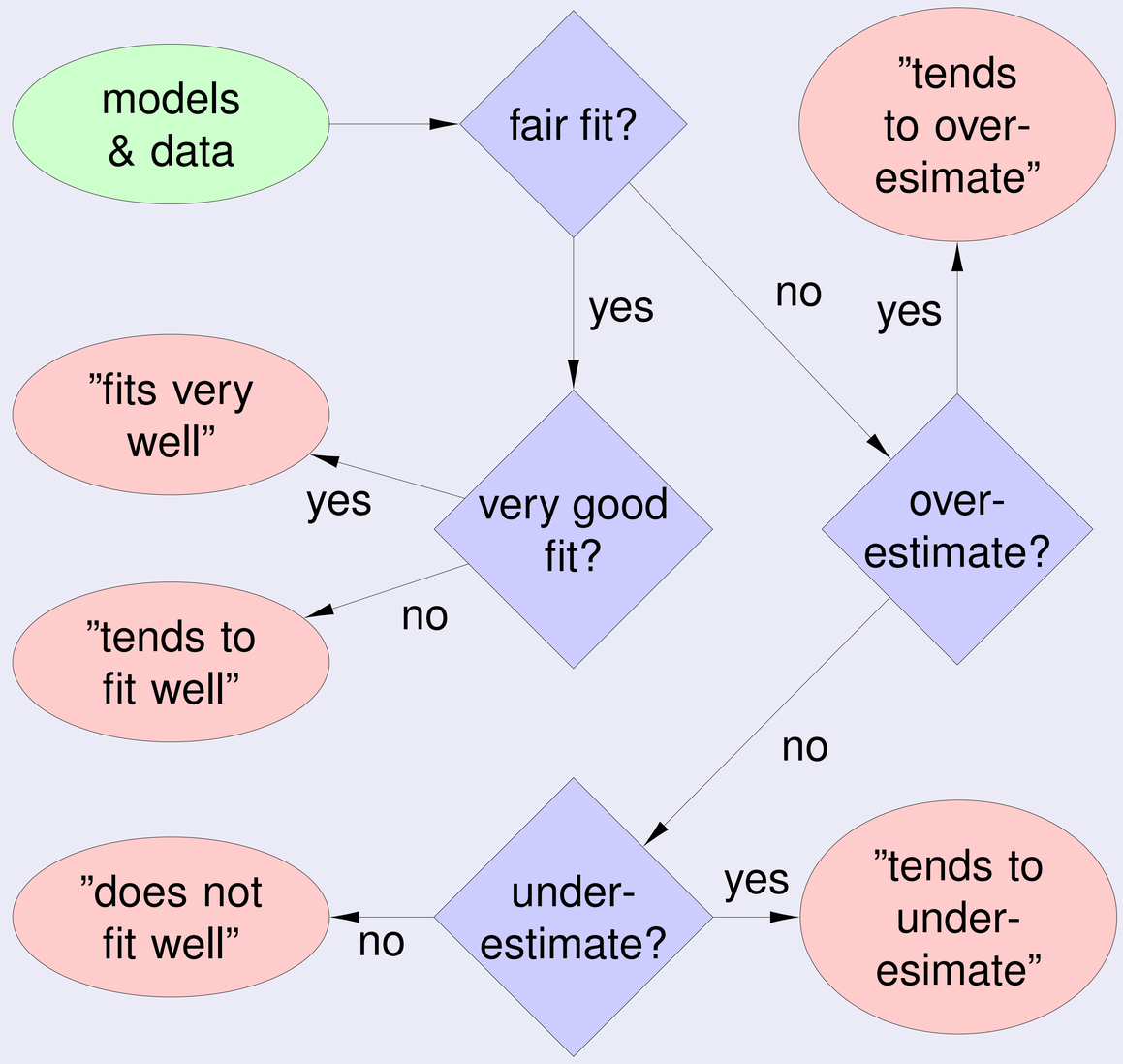
95% confidence intervals for predicted & observed data:

Solver	*n*	Predicted conf. intervals		Observed median running time	
		Poly. model	Exp. model	Point estimates	Conf. intervals
WalkSAT/SKC	600	[**0.054**, **0.081**]	[**0.067**, **0.104**]	0.056	[0.050, 0.070]
	700	[**0.083**, **0.146**]	[**0.137**, **0.264**]	0.121	[0.105, 0.145]
	800	[**0.122**, **0.238**]	[0.277, 0.664]	0.180	[0.132, 0.209]
	900	[**0.170**, **0.373**]	[0.565, 1.676]	0.267	[0.222, 0.323]
	1000	[**0.229**, **0.557**]	[1.151, 4.200]	0.385	[0.327, 0.461]

Graphical Results



Automatically Generated Interpretation



- fair fit: $>$ 70% challenge points or $>$ 70% of larger half of challenge points within predicted bootstrap intervals
- very good fit: $>$ 95% challenge points within predicted bootstrap intervals
- over-/under-estimate: $>$ 75% challenge points or $>$ 75% of larger half of challenge points below/above predicted bootstrap intervals

References

[1] Holger H. Hoos. A bootstrap approach to analysing the scaling of empirical run-time data with problem size. Technical report, TR-2009-16, Department of Computer Science, University of British Columbia, 2009.

[2] Holger H. Hoos and Thomas Stützle. On the empirical scaling of run-time for finding optimal solutions to the travelling salesman problem. *European Journal of Operational Research*, 238(1):87–94, 2014.

[3] Zongxu Mu and Holger H. Hoos. On the empirical time complexity of random 3-SAT at the phase transition. *IJCAI* 2015, to appear.