On the empirical scaling of running time of p15S363 for solving RUE Instances

Empirical Scaling Analyser

21st June 2016

1 Introduction

This is the automatically generated report on the empirical scaling of the running time of p15S363 for solving RUE Instances.

2 Methodology

For our scaling analysis, we considered the following parametric models:

- $EXP\left[a,b\right]\left(n\right) = a \times b^{x}$ (2-parameter EXP)
- $Poly\left[a,b\right]\left(n\right) = a \times x^{b}$ (2-parameter Poly)
- $SQRTEXP\left[a,b\right]\left(n\right) = a \times b^{\sqrt{x}}$ (2-parameter SQRTEXP)

Note that the approach could be easily extended to other scaling models. For fitting parametric scaling models to observed data, we used the non-linear least-squares Levenberg-Marquardt algorithm.

Models were fitted to performance observations in the form of medians of the distributions of running times over sets of instances for given n, the instance size. To assess the fit of a given scaling model to observed data, we used root-mean-square error (RMSE).

Closely following [2, 3], we computed 95% bootstrap confidence intervals for the performance predictions obtained from our scaling models, based on 1000 bootstrap samples per instance set and 1000 automatically fitted variants of each scaling model. In the following, we say that a scaling model is in- consistent with observed data if the bootstrap confidence interval for the observed data is disjoint from the bootstrap confidence interval for predicted running times; we say that a scaling model is strongly consistent with observed data, if the observed median is fully contained within the bootstrap confidence interval for predicted running times. Also, we define residue of a model at a given size as the observed point estimate less the predicated value.

3 Dataset Description

The dataset contains running times of the p15S363 algorithm solving 6 sets of instances of different sizes. We split the running times into two categories, support ($n \leq 2000$) and challenge ($n > 2000$). The details of the dataset can be found in Tables 1 and 2.

n	500	1000	1500	2000
# instances	100	100	100	100
# running times	100	100	100	100
mean	1.848	7.683	23.2	48.93
coefficient of variation	0.7054	0.7422	1.047	0.7628
Q(0.1)	1.451	5.087	10.31	18.79
Q(0.25)	1.519	5.292	11.09	21.42
median	1.61	5.581	12.65	29.78
Q(0.75)	1.685	6.243	27	57.92
Q(0.9)	1.79	13.45	47.91	104.5

Table 1: Details of the running time dataset used as support data for model fitting.

n	2500	3000
# instances	100	100
# running times	100	100
mean	∞	∞
coefficient of variation	N/A	N/A
Q(0.1)	28.7	46.02
Q(0.25)	32.01	70.67
median	41.9	135
Q(0.75)	76.26	249.2
Q(0.9)	155.7	∞

Table 2: Details of the running time dataset used as challenge data for model fitting.

		Model	RMSE (support)	RMSE (challenge)
p15S363	EXP. Model	$0.95094 \times 1.0017^{x}$	0.34618	30.319
	Poly. Model	$4.3589 \times 10^{-8} \times x^{2.675}$	0.84016	34.759
	SQRTEXP. Model	$\mathbf{0.06958 \times 1.145^{\sqrt{x}}}$	**0.39499**	**19.033**

Table 3: Fitted models of the medians of the running times and RMSE values (in CPU sec). The models yielding more accurate predictions (as per RMSEs on challenge data) are shown in boldface.

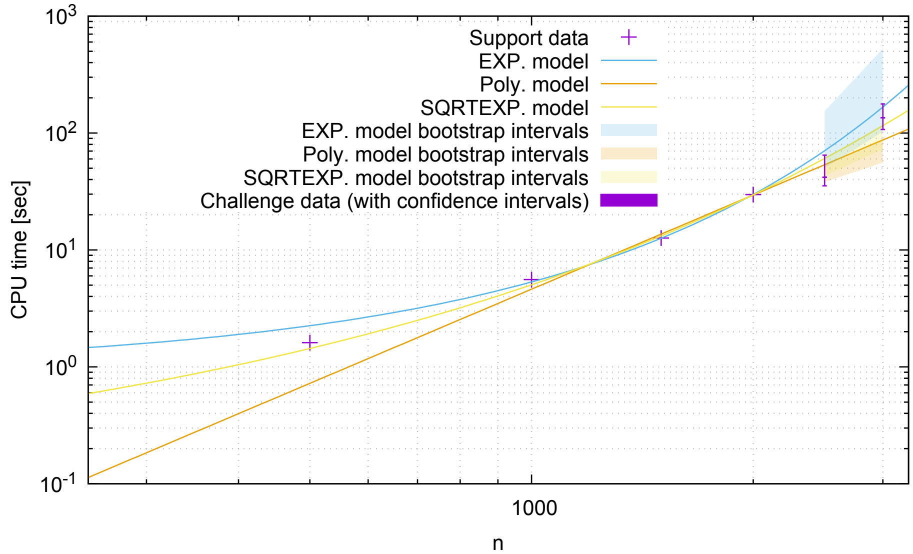

Figure 1: Fitted models of the medians of the running times. The models are fitted with the medians of the running times of p15S363 solving the set of RUE Instances of $500 \leq n \leq 2000$ variables, and are challenged by the medians of the running times of $2500 \leq n \leq 3000$ variables.

4 Empirical Scaling of Solver Performance

We first fitted our parametric scaling models to the medians of the running times of p15S363, as described in Section 2. The models were fitted using the medians of the running times for $500 \leq n \leq 2000$ (support) and later challenged with the medians of the running times for $2500 \leq n \leq 3000$. This resulted in the models, shown along with RMSEs on support and challenge data, shown in Table 3. In addition, we illustrate the fitted models of p15S363 in Figure 1, and the residues for the models in Figure 2.

But how much confidence should we have in these models? Are the RMSEs small enough that we should accept them? To answer this question, we assessed the fitted models using the bootstrap approach outlined in Section 2. Table 4 shows the bootstrap intervals of the model parameters, and Table 5 contains the bootstrap intervals for the support data. Challenging the models with extrapolation, as shown in Table 6, it is concluded that the EXP model tends to over-estimate the data, the Poly model under-estimates the data, and the SQRTEXP model under-estimates the data (as also illustrated in Figure 1).

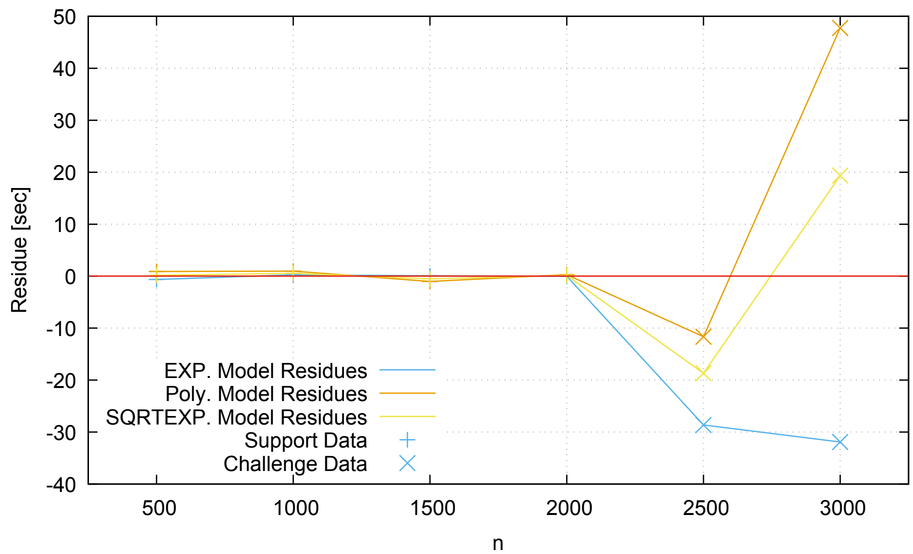

Figure 2: Residues of the fitted models of the medians of the running times.

Solver	Model	Confidence interval of a	Confidence interval of b
	EXP.	$[0.33168, 1.4922]$	$[1.0014, 1.0025]$
p15S363	Poly.	$[2.8428 \times 10^{-8}, 2.9333 \times 10^{-6}]$	$[2.0931, 2.7265]$
	SQRTEXP.	$[0.062519, 0.18669]$	$[1.1157, 1.1475]$

Table 4: 95% bootstrap intervals of model parameters for the medians of the running times

5 Conclusion

In this report, we presented an empirical analysis of the scaling behaviour of p15S363 on RUE Instances. We found the EXP model tends to over-estimate the data, the Poly model under-estimates the data, and the SQRTEXP model under-estimates the data.

References

[1] Jérémie Dubois-Lacoste, Holger H. Hoos, and Thomas Stützle. On the empirical scaling behaviour of state-of-the-art local search algorithms for the euclidean tsp. In *GECCO*. ACM, 2015.

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

[3] 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.

Solver	n	Predicted confidence intervals EXP. model	Observed median run-time Point estimates	Observed median run-time Confidence intervals
p15S363	500	$[\mathbf{1.132}, \mathbf{3.016}]$*	1.61	$[1.583, 1.646]$
	1000	$[\mathbf{3.859}, \mathbf{6.146}]$*	5.581	$[5.512, 5.726]$
	1500	$[\mathbf{11.83}, \mathbf{14.09}]$#	12.65	$[12.07, 14.35]$
	2000	$[\mathbf{24.54}, \mathbf{45.83}]$#	29.78	$[24.53, 45.94]$
Solver	n	**Predicted confidence intervals Poly. model**	**Point estimates**	**Confidence intervals**
p15S363	500	$[0.668, 1.359]$	1.61	$[1.583, 1.646]$
	1000	$[\mathbf{4.434}, \mathbf{5.82}]$*	5.581	$[5.512, 5.726]$
	1500	$[\mathbf{12.85}, \mathbf{14.34}]$	12.65	$[12.07, 14.35]$
	2000	$[\mathbf{23.75}, \mathbf{29.54}]$	29.78	$[24.53, 45.94]$
Solver	n	**Predicted confidence intervals SQRTEXP. model**	**Point estimates**	**Confidence intervals**
p15S363	500	$[\mathbf{1.36}, \mathbf{2.147}]$*	1.61	$[1.583, 1.646]$
	1000	$[\mathbf{4.864}, \mathbf{5.981}]$*	5.581	$[5.512, 5.726]$
	1500	$[\mathbf{12.32}, \mathbf{13.78}]$#	12.65	$[12.07, 14.35]$
	2000	$[\mathbf{23.96}, \mathbf{29.72}]$	29.78	$[24.53, 45.94]$

Table 5: 95% bootstrap confidence intervals for the medians of the running time predictions and observed running times on RUE Instances. The instance sizes shown here are those used for fitting the models. Bootstrap intervals on predictions that are weakly consistent with the observed point estimates are shown in boldface, those that are consistent are marked by plus signs (+), and those that fully contain the confidence intervals on observations are marked by asterisks (*).

Solver	n	Predicted confidence intervals EXP. model	Observed median run-time Point estimates	Observed median run-time Confidence intervals
p15S363	2500	$[\mathbf{50.14}, \mathbf{154}]$	41.9	$[35.35, 64.72]$
	3000	$[\mathbf{102.2}, \mathbf{525.1}]$*	135	$[107.2, 177.2]$
Solver	n	**Predicted confidence intervals Poly. model**	**Point estimates**	**Confidence intervals**
p15S363	2500	$[\mathbf{38.07}, \mathbf{54.13}]$#	41.9	$[35.35, 64.72]$
	3000	$[55.77, 89.02]$	135	$[107.2, 177.2]$
Solver	n	**Predicted confidence intervals SQRTEXP. model**	**Point estimates**	**Confidence intervals**
p15S363	2500	$[\mathbf{42.94}, \mathbf{61.24}]$	41.9	$[35.35, 64.72]$
	3000	$[\mathbf{72.4}, \mathbf{118.1}]$	135	$[107.2, 177.2]$

Table 6: 95% bootstrap confidence intervals for the medians of the running time predictions and observed running times on RUE Instances. The instance sizes shown here are larger than those used for fitting the models. Bootstrap intervals on predictions that are weakly consistent with the observed data are shown in boldface, those that are strongly consistent are marked by sharps (#), and those that fully contain the confidence intervals on observations are marked by asterisks (*).