On the empirical scaling of running time of LKH2 for solving RUE

Empirical Scaling Analyser

24th July 2017

1 Introduction

This is the automatically generated report on the empirical scaling of the running time of LKH2 for solving RUE.

2 Methodology

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

- $Poly\,[a, b]\,(n) = a \times x^b$ (2-parameter Poly)
- $Exp\,[a, b]\,(n) = a \times b^x$ (2-parameter Exp)
- $RootExp\,[a, b]\,(n) = a \times b^{\sqrt{x}}$ (2-parameter RootExp)
- $ExpLogRoot\,[a, b]\,(n) = a \times b^{\log(x)*\sqrt{x}}$ (2-parameter ExpLogRoot)
- $PolyL1\,[a]\,(n) = a \times x^{2.7155}$ (1-parameter PolyL1)
- $PolyU1\,[a]\,(n) = a \times x^{3.1281}$ (1-parameter PolyU1)

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 Q90s 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 LKH2 algorithm solving 15 sets of instances of different sizes. We split the running times into two categories, support ($n \leq 1500$) and challenge ($n > 1500$). The details of the dataset can be found in Tables 1 and 2.

n	500	600	700	800
# instances	1000	1000	1000	1000
# running times	1000	1000	1000	1000
mean	3.762	9.16	7.851	14.23
coefficient of variation	5.595	13.2	3.038	4.923
Q(0.1)	0.53	0.79	1.19	1.56
Q(0.25)	0.7	1.06	1.64	2.18
median	1.11	1.77	2.73	3.85
Q(0.75)	2.25	3.47	5.51	8.46
Q(0.9)	5.32	7.38	12.69	20.45

n	900	1000	1100	1200
# instances	1000	1000	1000	1000
# running times	1000	1000	1000	1000
mean	53.36	20.57	34.69	35.54
coefficient of variation	20.01	3.574	6.072	2.586
Q(0.1)	2.24	2.72	3.51	4.59
Q(0.25)	3.03	3.86	5.27	6.67
median	5.75	7.26	9.45	11.85
Q(0.75)	11.83	16.08	19.62	24.74
Q(0.9)	27.52	33.75	53.86	66.31

n	1300	1400	1500
# instances	1000	1000	1000
# running times	1000	1000	1000
mean	56.94	84.56	96.15
coefficient of variation	3.456	4.509	4.6
Q(0.1)	5.69	6.95	7.93
Q(0.25)	7.84	9.79	11.71
median	14.04	18.73	24.03
Q(0.75)	35.17	45	55.41
Q(0.9)	98.97	120.5	167.1

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

n	2000	2500	3000	3500
# instances	1000	100	100	100
# running times	1000	100	100	100
mean	234.9	742.7	611.9	5519
coefficient of variation	3.544	4.682	1.412	7.591
Q(0.1)	18.87	36.69	67.36	84.12
Q(0.25)	29.29	65.56	143.6	144.8
median	62.64	137	249.4	382.8
Q(0.75)	147.1	333.2	628.3	1414
Q(0.9)	391.3	790.8	1388	3265

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

		Model	RMSE (support)	RMSE (challenge)
LKH2	Poly. Model	$\mathbf{3.456 \times 10^{-10} \times x^{3.6753}}$	**3.6112**	**424.38**
	Exp. Model	1.7321×1.0031^{x}	3.3153	36735
	RootExp. Model	$0.047126 \times 1.2346^{\sqrt{x}}$	2.8221	4845.1
	ExpLogRoot. Model	$0.23483 \times 1.0234^{\log(x)*\sqrt{x}}$	2.8072	7212.9
	PolyL1. Model	$3.4649 \times 10^{-7} \times x^{2.7155}$	10.007	935.39
	PolyU1. Model	$1.7878 \times 10^{-8} \times x^{3.1281}$	6.1996	543.08

Table 3: Fitted models of the Q90s 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.

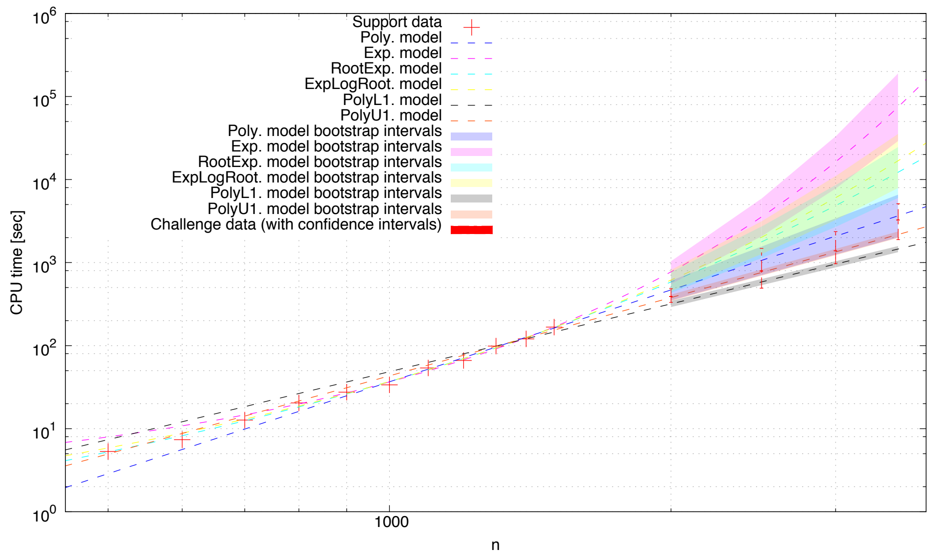

Figure 1: Fitted models of the Q90s of the running times. The models are fitted with the Q90s of the running times of LKH2 solving the set of RUE of $500 \leq n \leq 1500$ variables, and are challenged by the Q90s of the running times of $2000 \leq n \leq 3500$ variables.

4 Empirical Scaling of Solver Performance

We first fitted our parametric scaling models to the Q90s of the running times of LKH2, as described in Section 2. The models were fitted using the Q90s of the running times for $500 \leq n \leq 1500$ (support) and later challenged with the Q90s of the running times for $2000 \leq n \leq 3500$. 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 LKH2 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 Poly model fits the data very well, the Exp model over-estimates the data, the RootExp model over-estimates the data, the ExpLogRoot model over-estimates the data, the PolyL1 model under-estimates the data, and the PolyU1 model tends to

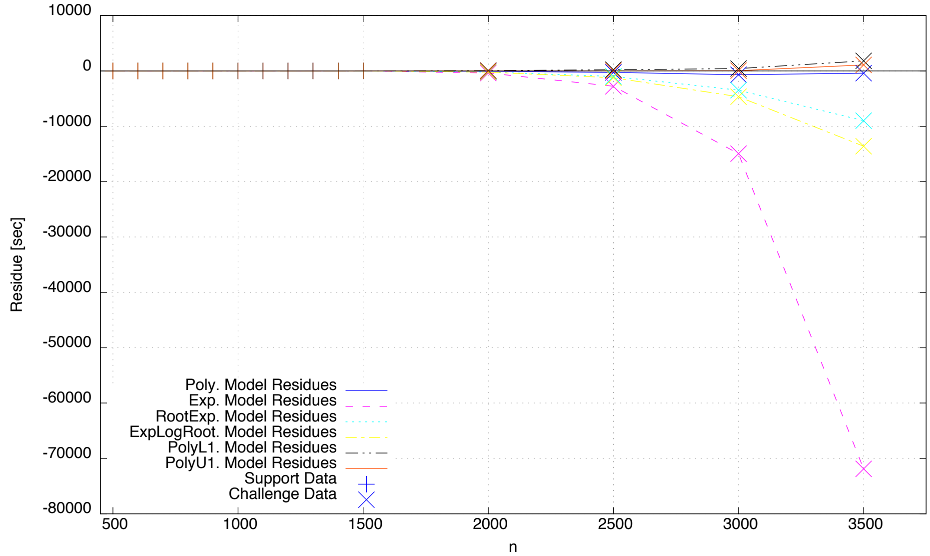

Figure 2: Residues of the fitted models of the Q90s of the running times.

Solver	Model	Confidence interval of a	Confidence interval of b
LKH2	Poly.	$[4.5581 \times 10^{-12}, 1.6362 \times 10^{-8}]$	$[3.1277, 4.2734]$
	Exp.	$[1.0016, 2.8495]$	$[1.0026, 1.0035]$
	RootExp.	$[0.01526, 0.1272]$	$[1.1994, 1.2732]$
	ExpLogRoot.	$[0.097364, 0.50786]$	$[1.0202, 1.0269]$
	PolyL1.	$[3.1674 \times 10^{-7}, 3.7584 \times 10^{-7}]$	
	PolyU1.	$[1.6279 \times 10^{-8}, 1.9436 \times 10^{-8}]$	

Table 4: 95% bootstrap intervals of model parameters for the Q90s of the running times

under-estimate the data (as also illustrated in Figure 1).

5 Conclusion

In this report, we presented an empirical analysis of the scaling behaviour of LKH2 on RUE. We found the Poly model fits the data very well, the Exp model over-estimates the data, the RootExp model over-estimates the data, the ExpLogRoot model over-estimates the data, the PolyL1 model under-estimates the data, and the PolyU1 model tends to under-estimate 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	Observed median run-time	
		Poly. model	Point estimates	Confidence intervals
LKH2	500	[**1.611**, **4.606**]	5.32	[4.34, 6.67]
	600	[**3.51**, **8.239**]*	7.38	[6.61, 8.18]
	700	[**6.772**, **13.34**]#	12.69	[11.05, 19.61]
	800	[**11.96**, **20.28**]	20.45	[17.07, 26.69]
	900	[**19.86**, **29.68**]#	27.52	[24.65, 32.26]
	1000	[**30.99**, **41.72**]#	33.75	[29.85, 41.79]
	1100	[**46.25**, **57.25**]#	53.86	[42.55, 65.97]
	1200	[**65.86**, **77.34**]#	66.31	[55.13, 77.25]
	1300	[**88.65**, **103.5**]#	98.97	[76.95, 120]
	1400	[**114.1**, **138.5**]#	120.5	[99.6, 141.5]
	1500	[**142.6**, **183.9**]#	167.1	[137.2, 194.9]

Solver	n	Predicted confidence intervals	Observed median run-time	
		Exp. model	Point estimates	Confidence intervals
LKH2	500	[**5.634**, **10.65**]	5.32	[4.34, 6.67]
	600	[**7.987**, **13.9**]	7.38	[6.61, 8.18]
	700	[**11.33**, **18.23**]#	12.69	[11.05, 19.61]
	800	[**16.05**, **23.85**]#	20.45	[17.07, 26.69]
	900	[**22.68**, **31.2**]#	27.52	[24.65, 32.26]
	1000	[**32**, **41.23**]#	33.75	[29.85, 41.79]
	1100	[**44.85**, **54.51**]#	53.86	[42.55, 65.97]
	1200	[**62.25**, **72.89**]#	66.31	[55.13, 77.25]
	1300	[**84.49**, **98.3**]	98.97	[76.95, 120]
	1400	[**112.1**, **135.8**]#	120.5	[99.6, 141.5]
	1500	[**147.8**, **189.1**]#	167.1	[137.2, 194.9]

Solver	n	Predicted confidence intervals	Observed median run-time	
		RootExp. model	Point estimates	Confidence intervals
LKH2	500	[**3.413**, **7.502**]*	5.32	[4.34, 6.67]
	600	[**5.717**, **11.02**]*	7.38	[6.61, 8.18]
	700	[**9.178**, **15.78**]#	12.69	[11.05, 19.61]
	800	[**14.31**, **22.23**]#	20.45	[17.07, 26.69]
	900	[**21.66**, **30.56**]#	27.52	[24.65, 32.26]
	1000	[**31.83**, **41.44**]#	33.75	[29.85, 41.79]
	1100	[**45.88**, **56**]#	53.86	[42.55, 65.97]
	1200	[**64.32**, **75.31**]#	66.31	[55.13, 77.25]
	1300	[**86.68**, **101.1**]#	98.97	[76.95, 120]
	1400	[**113.2**, **137.2**]#	120.5	[99.6, 141.5]
	1500	[**145.2**, **186.5**]#	167.1	[137.2, 194.9]

Solver	n	Predicted confidence intervals	Observed median run-time	
		ExpLogRoot. model	Point estimates	Confidence intervals
LKH2	500	[**3.886**, **8.227**]*	5.32	[4.34, 6.67]
	600	[**6.21**, **11.67**]*	7.38	[6.61, 8.18]
	700	[**9.651**, **16.37**]#	12.69	[11.05, 19.61]
	800	[**14.74**, **22.7**]#	20.45	[17.07, 26.69]
	900	[**21.96**, **30.77**]#	27.52	[24.65, 32.26]
	1000	[**31.89**, **41.41**]#	33.75	[29.85, 41.79]
	1100	[**45.66**, **55.74**]#	53.86	[42.55, 65.97]
	1200	[**63.86**, **74.8**]#	66.31	[55.13, 77.25]
	1300	[**86.22**, **100.6**]#	98.97	[76.95, 120]
	1400	[**113**, **136.9**]#	120.5	[99.6, 141.5]
	1500	[**145.8**, **187**]#	167.1	[137.2, 194.9]

Solver	n	Predicted confidence intervals	Observed median run-time	
		PolyL1. model	Point estimates	Confidence intervals
	500	[6.757, 8.018]	5.32	[4.34, 6.67]
	600	[11.09, 13.15]	7.38	[6.61, 8.18]

Solver	n	Predicted confidence intervals Poly. model	Observed median run-time Point estimates	 Confidence intervals
LKH2	2000	$[\mathbf{351.3}, \mathbf{607.3}]\#$	391.3	$[331.7, 484.8]$
	2500	$[\mathbf{707.2}, \mathbf{1563}]\#$	790.8	$[491.9, 1483]$
	3000	$[\mathbf{1251}, \mathbf{3408}]\#$	1388	$[970.5, 2363]$
	3500	$[\mathbf{2009}, \mathbf{6567}]\#$	3265	$[1888, 5117]$
Solver	n	Predicted confidence intervals Exp. model	Observed median run-time Point estimates	 Confidence intervals
LKH2	2000	$[553.2, 1040]$	391.3	$[331.7, 484.8]$
	2500	$[2070, 5888]$	790.8	$[491.9, 1483]$
	3000	$[7760, 3.309 \times 10^4]$	1388	$[970.5, 2363]$
	3500	$[2.917 \times 10^4, 1.892 \times 10^5]$	3265	$[1888, 5117]$
Solver	n	Predicted confidence intervals RootExp. model	Observed median run-time Point estimates	 Confidence intervals
LKH2	2000	$[\mathbf{431}, \mathbf{768.1}]$	391.3	$[331.7, 484.8]$
	2500	$[\mathbf{1129}, \mathbf{2716}]$	790.8	$[491.9, 1483]$
	3000	$[2686, 8571]$	1388	$[970.5, 2363]$
	3500	$[5993, 2.463 \times 10^4]$	3265	$[1888, 5117]$
Solver	n	Predicted confidence intervals ExpLogRoot. model	Observed median run-time Point estimates	 Confidence intervals
LKH2	2000	$[\mathbf{452.3}, \mathbf{815.8}]$	391.3	$[331.7, 484.8]$
	2500	$[\mathbf{1264}, \mathbf{3148}]$	790.8	$[491.9, 1483]$
	3000	$[3265, 1.099 \times 10^4]$	1388	$[970.5, 2363]$
	3500	$[7901, 3.545 \times 10^4]$	3265	$[1888, 5117]$
Solver	n	Predicted confidence intervals PolyL1. model	Observed median run-time Point estimates	 Confidence intervals
LKH2	2000	$[\mathbf{291.5}, \mathbf{345.9}]$	391.3	$[331.7, 484.8]$
	2500	$[\mathbf{534.3}, \mathbf{634}]$	790.8	$[491.9, 1483]$
	3000	$[\mathbf{876.6}, \mathbf{1040}]$	1388	$[970.5, 2363]$
	3500	$[1332, 1581]$	3265	$[1888, 5117]$
Solver	n	Predicted confidence intervals PolyU1. model	Observed median run-time Point estimates	 Confidence intervals
LKH2	2000	$[\mathbf{344.8}, \mathbf{411.7}]\#$	391.3	$[331.7, 484.8]$
	2500	$[\mathbf{693}, \mathbf{827.4}]\#$	790.8	$[491.9, 1483]$
	3000	$[\mathbf{1226}, \mathbf{1464}]\#$	1388	$[970.5, 2363]$
	3500	$[\mathbf{1985}, \mathbf{2370}]$	3265	$[1888, 5117]$

Table 6: 95% bootstrap confidence intervals for the Q90s of the running time predictions and observed running times on RUE. 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 (*).