On the empirical scaling of running time of WalkSAT/SKC for solving random 3-SAT instances at the phase transition

Empirical Scaling Analyser

18th July 2017

1 Introduction

This is the automatically generated report on the empirical scaling of the running time of WalkSAT/SKC for solving random 3-SAT instances at the phase transition.

2 Methodology

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

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

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 100 bootstrap samples per instance set and 100 automatically fitted variants of each scaling model. To extend this idea, we calculated support and challenge RMSEs for each of the fitted models' predictions and the corresponding bootstrap samples of the observed data. We used these bootstrap sample RMSEs to calculate median and 95% confidence intervals of the support and challenge RMSEs for each model.

In the following, we say that a scaling model prediction is in-consistent with observed data if the bootstrap confidence interval for the observed data is disjoint from the bootstrap confidence interval for the predicted median running times; we say that a scaling model prediction is weakly consistent with the observed data if the bootstrap confidence interval for the prediction overlaps with the bootstrap confidence interval for the observed data; and, we say that a scaling model is strongly consistent with observed data, if the bootstrap confidence interval for the observed median is fully contained within the bootstrap confidence interval for predicted running times. Also, we define the residue of a model at a given size as the observed point estimate less the predicated value.

n	200	250	300	350
# instances	601	589	633	558
# running times	601	589	633	558
mean	0.006527	0.01671	0.04785	0.07433
coefficient of variation	1.932	2.708	7.148	4.636
Q(0.1)	0.000584	0.001108	0.001648	0.002231
Q(0.25)	0.000994	0.001882	0.003173	0.004306
median	0.002093	0.004457	0.007494	0.01094
Q(0.75)	0.005678	0.01213	0.02102	0.02984
Q(0.9)	0.01572	0.03676	0.05997	0.0899

n	400	450	500
# instances	579	572	578
# running times	579	572	578
mean	0.2162	0.2634	2.171
coefficient of variation	8.165	6.233	17.97
Q(0.1)	0.003448	0.005009	0.006445
Q(0.25)	0.007598	0.01004	0.01438
median	0.01825	0.02414	0.03651
Q(0.75)	0.05361	0.08692	0.1295
Q(0.9)	0.2451	0.3535	0.4501

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

3 Dataset Description

The dataset contains running times of the WalkSAT/SKC algorithm solving 12 sets of instances of different sizes . We split the running times into two categories, support ($n \leq 500$) and challenge ($n > 500$). The details of the dataset can be found in Tables 1 and 2.

4 Empirical Scaling of Solver Performance

We first fitted our parametric scaling models to the medians of the running times of WalkSAT/SKC, as described in Section 2. The models were fitted using the medians of the running times for $200 \leq n \leq 500$ (support) and later challenged with the medians of the running times for $600 \leq n \leq 1000$. 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 WalkSAT/SKC 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, Table 5 shows the bootstrap intervals of the model prediction RMSEs, and Table 6 contains the bootstrap intervals for the support data. Challenging the models with extrapolation, as shown in Table 7, it is concluded that the Exp model over-estimates the data, the RootExp model tends to fit the data, and the Poly model fits the data very well (as also illustrated in Figure 1). We base these statements on an analysis of the fraction of predicted bootstrap intervals that are strongly consistent, weakly consistent and disjoint from the observed bootstrap intervals for the challenge data. To provide stronger emphasis for the largest instance sizes, we also consider these fractions for the largest half of

n	600	700	800	900	1000
# instances	572	636	584	592	593
# running times	572	636	584	592	593
mean	2.503	3.303	2.772	15.54	30.16
coefficient of variation	13.32	7.855	5.129	6.333	5.432
Q(0.1)	0.01243	0.01829	0.02679	0.03593	0.05379
Q(0.25)	0.02394	0.03952	0.0549	0.07982	0.119
median	0.05641	0.1083	0.1797	0.2668	0.3845
Q(0.75)	0.2017	0.4775	0.7492	1.336	1.826
Q(0.9)	1.081	2.035	3.509	8.328	14.56

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

		Model	RMSE (support)	RMSE (challenge)
WalkSAT/SKC	Exp. Model	$0.00068923 \times 1.008^{x}$	5.137×10^{-6}	0.76018
	RootExp. Model	$2.8818 \times 10^{-5} \times 1.376^{\sqrt{x}}$	2.9317×10^{-6}	0.15778
	Poly. Model	$\mathbf{8.8435 \times 10^{-11} \times x^{3.1891}}$	**0.010305**	**0.031434**

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

the challenge instance sizes. To be precise, we say a model tends to fit the data if 90% or more of the predicted bootstrap intervals (or the larger half of the predicted intervals) are weakly consistent with the observed data; we say a model over-estimates the data if more than 70% of the predicted bootstrap intervals (or more than 70% of the larger half of the predicted bootstrap intervals) are disjoint from the observed bootstrap intervals and are above the observed intervals; and we say a model fits the data very well if 90% or more of the predicted bootstrap intervals (or the larger half of the predicted intervals) are strongly consistent with the observed data and at least 90% of the predicted bootstrap intervals are weakly consistent with the observed data.

5 Conclusion

In this report, we presented an empirical analysis of the scaling behaviour of WalkSAT/SKC on random 3-SAT instances at the phase transition. We found the Exp model over-estimates the data, the RootExp model tends to fit the data, and the Poly model fits the data very well.

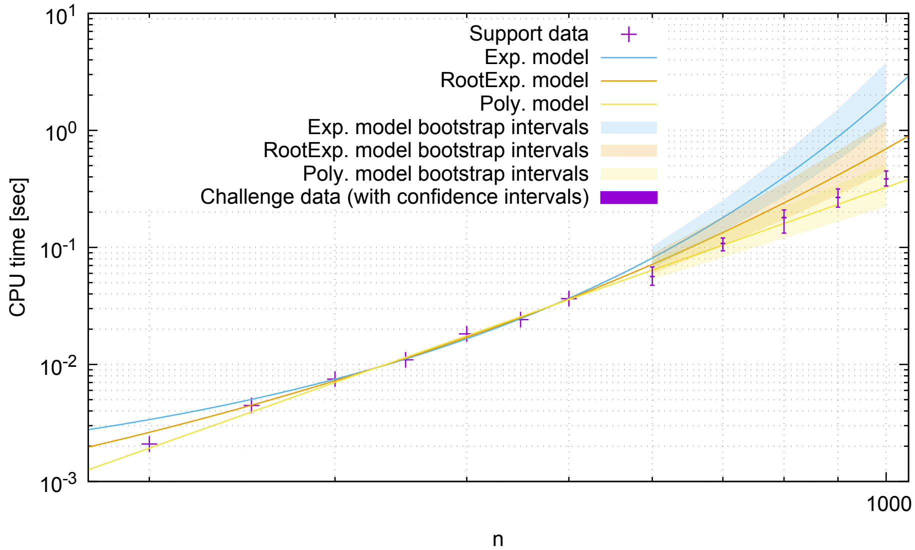

Figure 1: Fitted models of the medians of the running times. The models are fitted with the medians of the running times of WalkSAT/SKC solving the set of random 3-SAT instances at the phase transition of $200 \leq n \leq 500$ variables, and are challenged by the medians of the running times of $600 \leq n \leq 1000$ variables.

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 *Proceedings of the 17th Genetic and Evolutionary Computation Conference*, (GECCO '15), pages 377–384, 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, Department of Computer Science, 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.

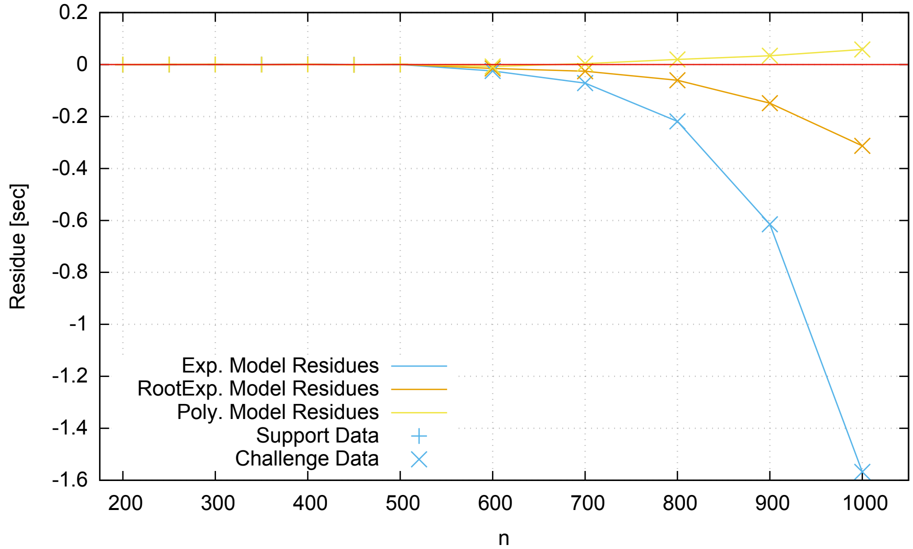

Figure 2: Residues of the fitted models of the medians of the running times.

Solver	Model	Confidence interval of a	Confidence interval of b
	Exp.	$[0.0004476, 0.0010113]$	$[1.007, 1.0091]$
WalkSAT/SKC	RootExp.	$[1.0905 \times 10^{-5}, 6.1897 \times 10^{-5}]$	$[1.3243, 1.4433]$
	Poly.	$[4.5554 \times 10^{-12}, 1.011 \times 10^{-9}]$	$[2.7812, 3.6834]$

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

Solver	Model	Support RMSE		Challenge RMSE	
		Median	Confidence Interval	Median	Confidence Interval
	Exp.	0.0013414	$[0.00066148, 0.0023947]$	0.76758	$[0.34781, 1.633]$
WalkSAT/SKC	RootExp.	0.0010909	$[0.00044449, 0.0025122]$	0.16244	$[0.026794, 0.4174]$
	Poly.	**0.001276**	$[\mathbf{0.00050239}, \mathbf{0.0028019}]$	**0.041205**	$[\mathbf{0.01038}, \mathbf{0.09576}]$

:

Table 5: Median and 95% bootstrap intervals of model prediction RMSEs for the medians of the running times. The models yielding the most accurate predictions (as per median challenge RMSE) are shown in boldface.

Solver	n	Predicted confidence intervals Exp. model	Observed median run-time Point estimates	 Confidence intervals
WalkSAT/SKC	200	$[0.002754, 0.004125]$	0.002093	$[0.001899, 0.002539]$
	250	$[\mathbf{0.004279}, \mathbf{0.005926}]$	0.004457	$[0.003664, 0.005377]$
	300	$[\mathbf{0.006648}, \mathbf{0.008532}]$*	0.007494	$[0.006795, 0.008479]$
	350	$[\mathbf{0.01034}, \mathbf{0.01227}]$	0.01094	$[0.009792, 0.01258]$
	400	$[\mathbf{0.01563}, \mathbf{0.01822}]$	0.01825	$[0.01601, 0.0201]$
	450	$[\mathbf{0.0228}, \mathbf{0.02791}]$	0.02414	$[0.02149, 0.02995]$
	500	$[\mathbf{0.03263}, \mathbf{0.04218}]$	0.03651	$[0.03177, 0.04233]$
Solver	**n**	**Predicted confidence intervals RootExp. model**	**Observed median run-time Point estimates**	** Confidence intervals**
WalkSAT/SKC	200	$[\mathbf{0.002052}, \mathbf{0.00334}]$	0.002093	$[0.001899, 0.002539]$
	250	$[\mathbf{0.003717}, \mathbf{0.005335}]$	0.004457	$[0.003664, 0.005377]$
	300	$[\mathbf{0.006362}, \mathbf{0.008301}]$	0.007494	$[0.006795, 0.008479]$
	350	$[\mathbf{0.01049}, \mathbf{0.01242}]$	0.01094	$[0.009792, 0.01258]$
	400	$[\mathbf{0.01612}, \mathbf{0.01876}]$	0.01825	$[0.01601, 0.0201]$
	450	$[\mathbf{0.02324}, \mathbf{0.02848}]$	0.02414	$[0.02149, 0.02995]$
	500	$[\mathbf{0.03229}, \mathbf{0.0418}]$	0.03651	$[0.03177, 0.04233]$
Solver	**n**	**Predicted confidence intervals Poly. model**	**Observed median run-time Point estimates**	** Confidence intervals**
WalkSAT/SKC	200	$[\mathbf{0.001369}, \mathbf{0.002609}]$*	0.002093	$[0.001899, 0.002539]$
	250	$[\mathbf{0.003143}, \mathbf{0.004819}]$	0.004457	$[0.003664, 0.005377]$
	300	$[\mathbf{0.006063}, \mathbf{0.00811}]$	0.007494	$[0.006795, 0.008479]$
	350	$[\mathbf{0.01061}, \mathbf{0.01266}]$	0.01094	$[0.009792, 0.01258]$
	400	$[\mathbf{0.01661}, \mathbf{0.01927}]$	0.01825	$[0.01601, 0.0201]$
	450	$[\mathbf{0.02363}, \mathbf{0.02901}]$	0.02414	$[0.02149, 0.02995]$
	500	$[\mathbf{0.03187}, \mathbf{0.04135}]$	0.03651	$[0.03177, 0.04233]$

Table 6: 95% bootstrap confidence intervals for the medians of the running time predictions and observed running times on random 3-SAT instances at the phase transition. 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 and those that are strongly consistent are marked by asterisks (*).

Solver	n	Predicted confidence intervals Exp. model	Observed median run-time Point estimates	 Confidence intervals
WalkSAT/SKC	600	$[\mathbf{0.06649}, \mathbf{0.1016}]$	0.05641	$[0.04749, 0.06806]$
	700	$[0.1347, 0.2526]$	0.1083	$[0.09343, 0.1206]$
	800	$[0.2724, 0.628]$	0.1797	$[0.1324, 0.2091]$
	900	$[0.5513, 1.522]$	0.2668	$[0.2204, 0.3156]$
	1000	$[1.108, 3.769]$	0.3845	$[0.3348, 0.4494]$
Solver	**n**	**Predicted confidence intervals RootExp. model**	**Observed median run-time Point estimates**	** Confidence intervals**
WalkSAT/SKC	600	$[\mathbf{0.05936}, \mathbf{0.08853}]$	0.05641	$[0.04749, 0.06806]$
	700	$[\mathbf{0.1035}, \mathbf{0.1822}]$	0.1083	$[0.09343, 0.1206]$
	800	$[\mathbf{0.1734}, \mathbf{0.3566}]$	0.1797	$[0.1324, 0.2091]$
	900	$[\mathbf{0.2809}, \mathbf{0.6702}]$	0.2668	$[0.2204, 0.3156]$
	1000	$[\mathbf{0.4416}, \mathbf{1.198}]$	0.3845	$[0.3348, 0.4494]$
Solver	**n**	**Predicted confidence intervals Poly. model**	**Observed median run-time Point estimates**	** Confidence intervals**
WalkSAT/SKC	600	$[\mathbf{0.05335}, \mathbf{0.07789}]$	0.05641	$[0.04749, 0.06806]$
	700	$[\mathbf{0.0823}, \mathbf{0.1374}]$*	0.1083	$[0.09343, 0.1206]$
	800	$[\mathbf{0.1196}, \mathbf{0.2247}]$*	0.1797	$[0.1324, 0.2091]$
	900	$[\mathbf{0.1663}, \mathbf{0.3468}]$*	0.2668	$[0.2204, 0.3156]$
	1000	$[\mathbf{0.223}, \mathbf{0.5113}]$*	0.3845	$[0.3348, 0.4494]$

Table 7: 95% bootstrap confidence intervals for the medians of the running time predictions and observed running times on random 3-SAT instances at the phase transition. 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 and those that are strongly consistent are marked by asterisks (*).