Detailed SATzilla Results from the Data Analysis Track of the 2011 SAT Competition

Experimental setup

Solvers: For each category, we constructed portfolios and oracles (SATzilla with perfect selector) using all non-portfolio, non-parallel solvers from Phase 2 of the competition as component solvers. Note that the "virtual best solver" (VBS) includes all solvers from Phase 2. Table 1 gives the total number of solvers used for each category.

Data Set	Application	Crafted	Random
Number of all solvers (used in VBS)	31	25	17
Number of candidates (used in SATzilla and Oracle)	18	15	9

Table 1. The number of solvers from Phase 2 of the 2011 SAT Competition; the number of solvers (excluding portfolio-based and multi-threaded procedures) used for building SATzilla 2011 (and Oracle). See attachments for the names of the solvers.

Features We used 115 features similar to those used by SATzilla in the 2009 SAT Competition, which fall into 9 categories: problem size, variable graph, clause graph, variable-clause graph, balance, proximity to Horn formula, local search probing, clause learning, and survey propagation. Feature computation times averaged 31.4 CPU seconds (*Random*), 51.8 CPU seconds (*Crafted*), and 158.5 CPU seconds (*Application*).

Methods We used the SATzilla procedure (based on cost-sensitive classification models) described in *Hydra-MIP: Automated Algorithm Configuration and Selection for Mixed Integer Programming,* RCRA workshop, IJCAI 2011. We used 10-fold cross validation to obtain an unbiased estimate of SATzilla's performance. First, we eliminated all instances that could not be solved by any candidate solver *(we denote those instances as U)*. Then, we performed 10-fold cross validation as follows. We partitioned the remaining instances *(denoted S)* into 10 disjoint sets. Setting aside each of these sets in turn as the test set, we constructed SATzilla with the other 9 sets of instances as training data. Finally we measured SATzilla's runtime on the test set. In the end, we report SATzilla's average performance across the 10 test sets. Since none of the candidate solvers can solve any instance in *U,* SATzilla is also guaranteed to time out on *U*.

Data Runtime data was provided by the organisers of the 2011 SAT competition. All feature

computations were performed by *Daniel Le Berre* on a quad core computer with 4GB of RAM and running Linux, using our code. Four instances (from the *Crafted* category) out of 1200 had no feature values due to a database problem caused by duplicated file names[1]. For the time being, we have treated these as timeouts for SATzilla to obtain a lower bound on SATzilla's performance.

Results

Application:

Solver	Virtual best solver	Oracle	SATzilla 2011 (Application)	Gold medalist: Glucose 2
Average Runtime (CPU seconds)	1104	1138	1707	1856
Solved Percentage	84.7%	84.3%	75.0%	71.7%

Table 2: Comparison of SATzilla 2011 (Application) with virtual best solver, Oracle (SATzilla with perfect selector), and the gold medalist (Glucose 2) on Application;
Timeout runs are counted as the cutoff of 5000 CPU seconds.

---

[1] Four pairs of instances in the *Crafted* category had identical names. Since the database for recording feature values used file name as the primary key, the feature values for four instances could not be written into the database. We sent a request to *Daniel Le Berre* for re-collecting these feature values, but have not yet received them.
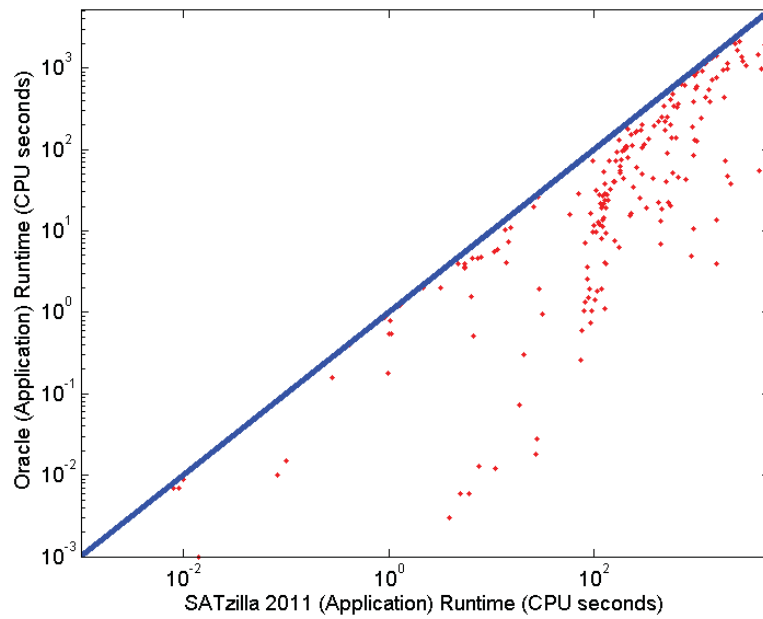
Figure 1: Comparison of SATzilla 2011 (Application) and Oracle (SATzilla with perfect selector) on Application; each point corresponds to one SAT instance.
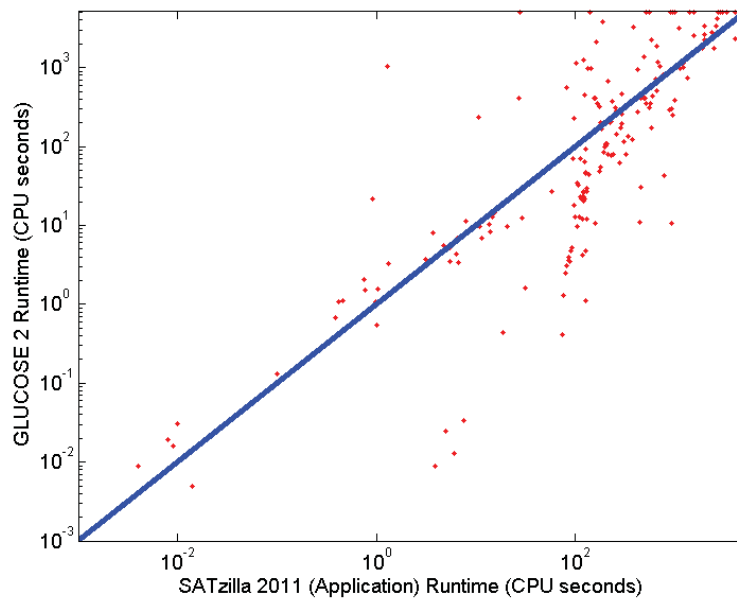

Figure 2: Comparison of SATzilla 2011 (Application) and the gold medalist (Glucose 2) on Application; each point corresponds to one SAT instance.
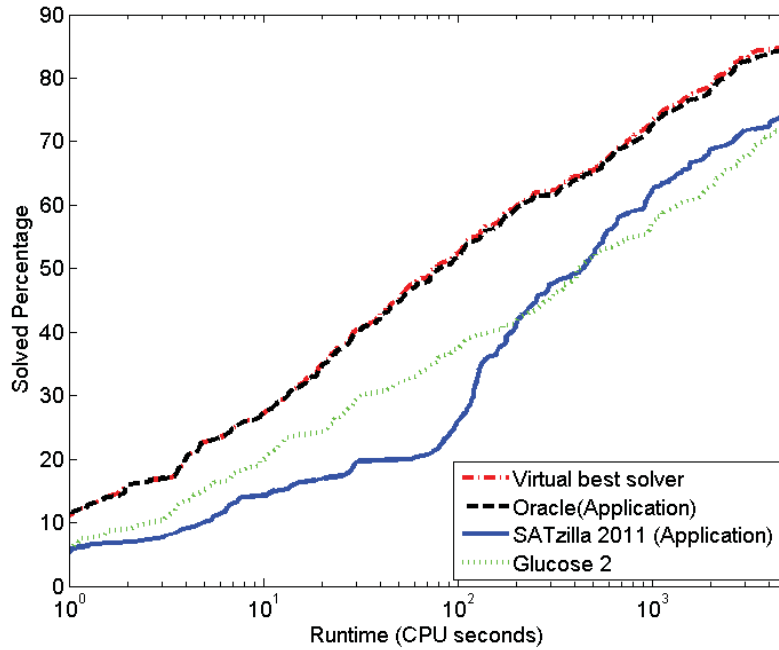
Figure 3: Runtime CDF for SATzilla 2011 (Application), virtual best solver, Oracle (SATzilla with perfect selector), and the gold medalist (Glucose 2) on Application.

Crafted:

Solver	Virtual best solver	Oracle	SATzilla 2011 (Crafted)	Gold medalist: 3S
Average Runtime (CPU seconds)	1542	1667	2155	2602
Solved Percentage	76.3%	73.7%	64.7%	54.3%

Table 3: Comparison of SATzilla 2011 (Crafted) with virtual best solver, Oracle (SATzilla with perfect selector), and the gold medalist (3S) on Crafted;
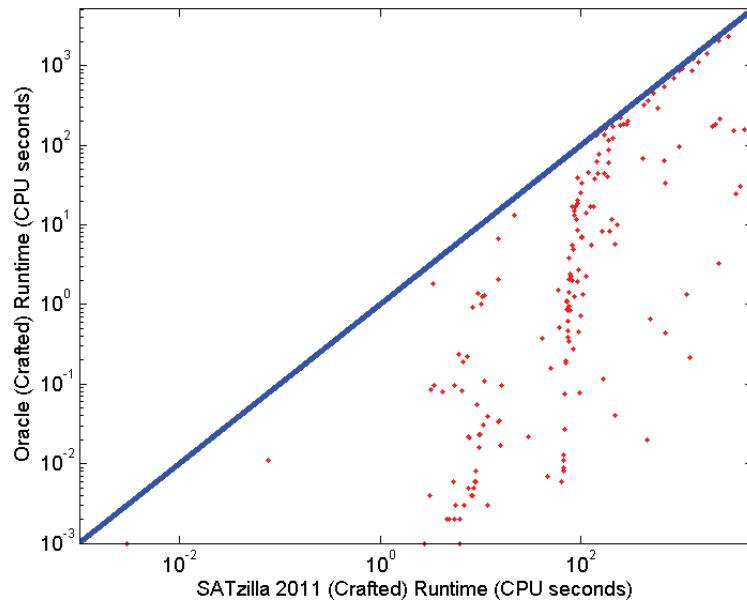*Timed-out runs are counted as the cutoff of 5000 CPU seconds.*

Figure 4: Comparison of SATzilla 2011 (Crafted) and Oracle (SATzilla with perfect selector) on Crafted; each point corresponds to one SAT instance.
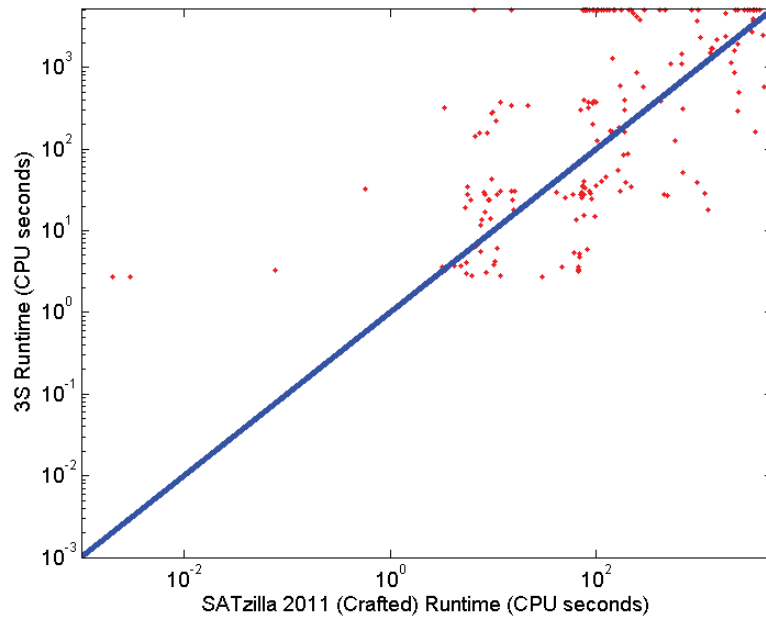

Figure 5: Comparison of SATzilla 2011 (Crafted) and the gold medalist (3S) on Crafted; each point corresponds to one SAT instance.
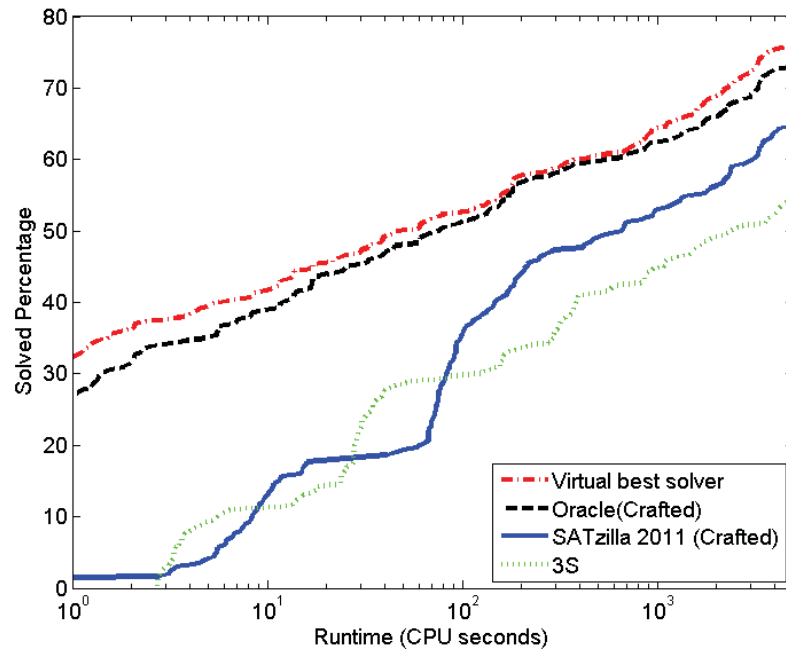
Figure 6: Runtime CDF for SATzilla 2011 (Crafted), virtual best solver, Oracle (SATzilla with perfect selector), and the gold medalist (3S) on Crafted

Random:

Solver	Virtual best solver	Oracle	SATzilla 2011 (Random)	Gold medalist: 3S
Average Runtime (CPU seconds)	1074	1087	1313	1836
Solved Percentage	82.2%	82.0%	79.2%	68.0%

Table 4: Comparison of SATzilla 2011 (Random) with virtual best solver, Oracle (SATzilla with perfect selector), and the gold medalist (3S) on Random;
Timed-out runs are counted as the cutoff of 5000 CPU seconds
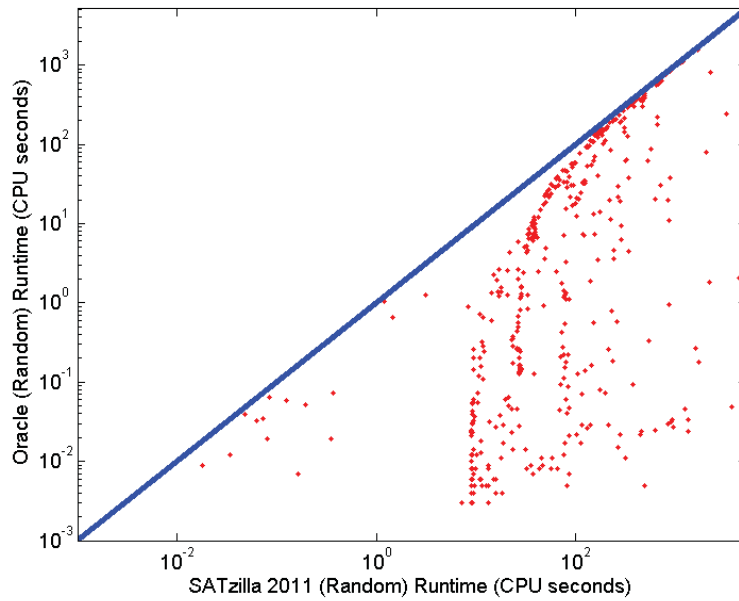
Figure 7: Comparison of SATzilla 2011 (Random) and Oracle (SATzilla with perfect selector) on Random; each point corresponds to one SAT instance.
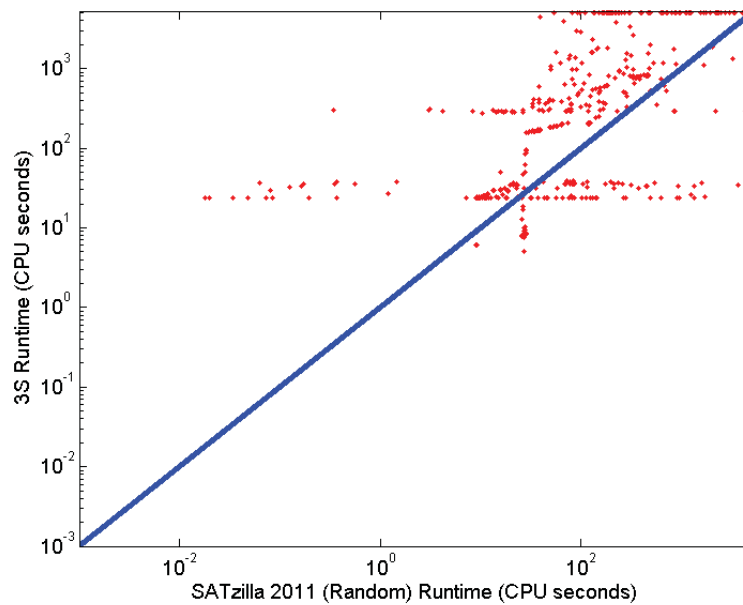

Figure 8: Comparison of SATzilla 2011 (Random) and the gold medalist (3S) on Random; each point corresponds to one SAT instance.
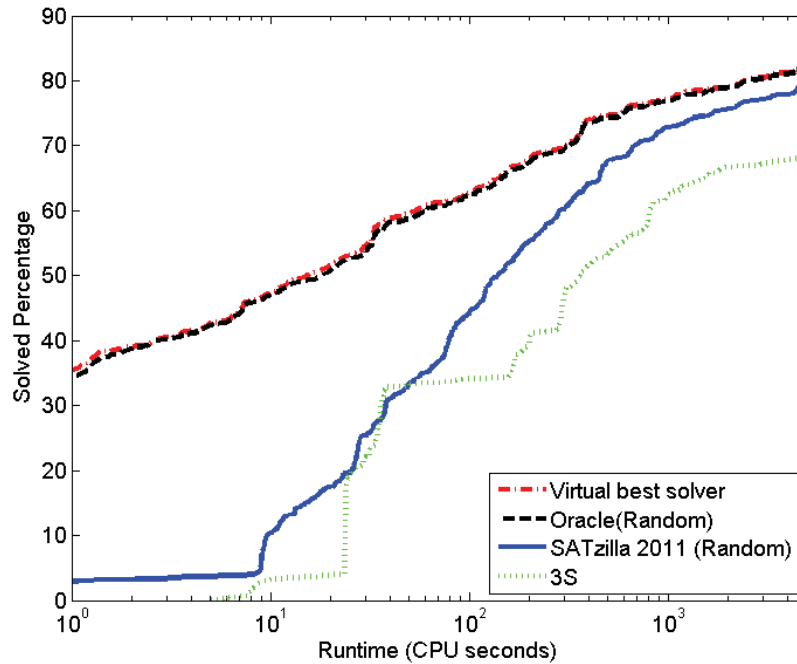
Figure 9: Runtime CDF for SATzilla 2011 (Random), virtual best solver, Oracle (SATzilla with perfect selector), and the gold medalist (3S) on Random.

Attachments:

1. Solvers used for building virtual best solver and SATzilla 2011 (6 files)

File "*[data set]_vb_solvers.txt*" lists all the solvers used for building virtual best solver. File "*[data set]_zilla_solvers.txt*" lists all the candidate solvers for building SATzilla 2011 and Oracle. Here, *[data set]* is chosen from "*application*", "*crafted*", and "*random*".

2. Runtime data for virtual best solver, Oracle, SATzilla 2011, and the gold medal winner (3 files)

File "*[data set]_runtime.csv*" contains runtime data for virtual best solver, Oracle, SATzilla 2011, and the gold medal winner. Here, *[data set]* is chosen from "*application*", "*crafted*", and "*random*".