Sparkle SAT Challenge 2018

Chuan Luo and Holger H. Hoos

LIACS
Universiteit Leiden
The Netherlands

FLoC 2018, Oxford, UK
12 July 2018

The state of the art in solving X ...

- ... is not defined by a single solver / solver configuration
- ... requires use of / interplay between
  multiple heuristic mechanisms
- ... has been substantially advanced by machine learning

Competitions ...

- ... have helped advance the state of the art in many fields (SAT, AI planning, machine learning, ...)
- ... are mostly focussed on single solvers, broad-spectrum performance
- ... often don't reflect state of the art
- ... typically don't provide effective incentive to improve state of the art

A new kind of competition:

- solvers submitted to competition platform
- state-of-the-art per-instande selector built based on all solvers
- solver contributions to overall performance assessed based on (relative) marginal contribution

  (Xu, Hutter, HH, Leyton-Brown 2012; Luo & Hoos – this event)
- full credit for contributions to selector performance goes to component solver authors

$\rightsquigarrow$ Sparkle (Luo & Hoos – this event)


1
2
3


XXXI Olimpíada Rio 2016


XXXI Olimpíada Rio 2016

Sparkle SAT Challenge 2018

- part of FLoC Olympic Games, coordinated with 2018 SAT Competition
- launched March 2018, leader board phase 5–15 April, final results now!
- 19 open-source solvers submitted, 4 hors-concours solvers included
- website: http://ada.liacs.nl/events/sparkle-sat-18

Training and testing sets

- training set: 1356 instances from 25 families,
  all solved instances (satisfiable + unsatisfiable) from main, application, crafted/hard-combinatorial tracks of 2014–2017 SAT Competitions + 2015 SAT Race
- testing set: 400 instances from 23 families,
  identical to testing set of main track of 2018 SAT Competition

Constructing the per-instance selector

- training set: 1356 instances from 25 families
- split training set into *core training set* and *validating set*
  - randomly select 15 instance families $\rightarrow$ core training set
  - remaining 10 families $\rightarrow$ validating set
- core training set: 893 instances from 15 families
- validating set: 463 instances from 10 families

- run AutoFolio (Lindauer *et al.* 2015) 100 times to obtain 100 per-instance selectors
  - train on core training set
  - choose selector with smallest PAR2 score on validating set

AutoFolio

- automatically configure flexible selector framework to find state-of-the-art, customised selectors

  (Lindauer, Hoos, Hutter, Schaub 2015)

- based on well-known, flexible per-instance algorithm selection framework: claspfolio 2 (Hoos & Lindauer & Schaub 2014)

- leverages state-of-the-art, general-purpose algorithm configurator: SMAC (Hutter, Hoos, Leyton-Brown 2011)

⇝ cutting-edge, robust algorithm selector construction in Sparkle

Assessing solver contributions

Given: set of solvers S; per-instance selector P based on S; instance set I

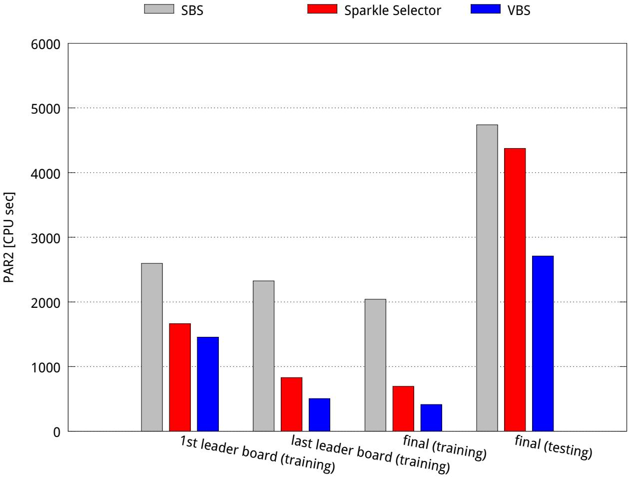
absolute marginal contribution (amc) of solver s on I:

$$amc(s, I) = \begin{cases} log_{10} \frac{PAR2(P \setminus \{s\}, I)}{PAR2(P, I)} & PAR2(P \setminus \{s\}, I) > PAR2(P, I) \\ 0 & \text{else} \end{cases}$$

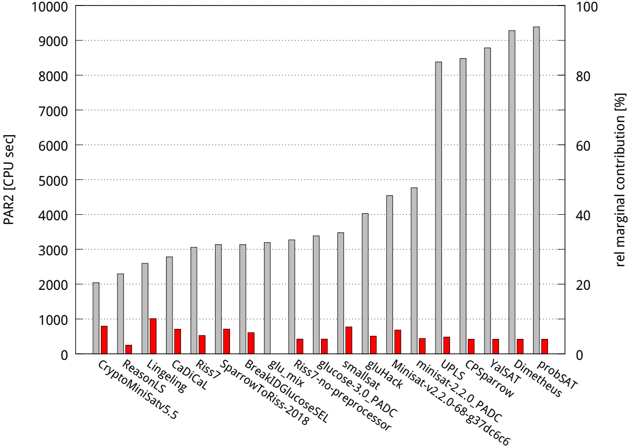
relative marginal contribution (rmc) of solver s of I:

$$rmc(s, I) = \frac{amc(s)}{\sum_{s' \in S} amc(s')}$$

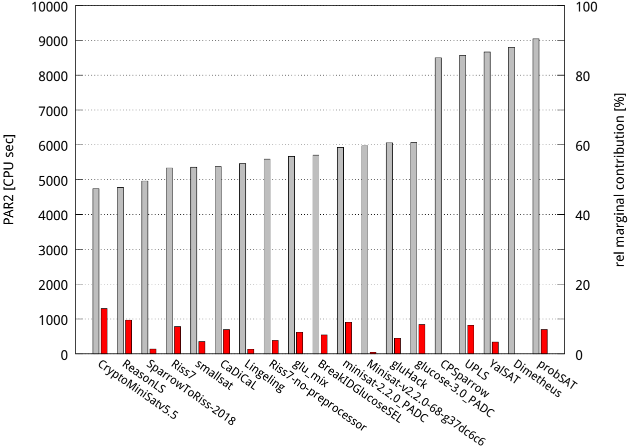
Improvement over time without hors-concours solvers



Stand-alone and relative marginal contribution on training set, without hors-concours solvers



Stand-alone and relative marginal contribution on testing set, without hors-concours solvers



Final results without hors-concours solvers, on testing set

PAR2 for SBS, VBS and Sparkle Selector

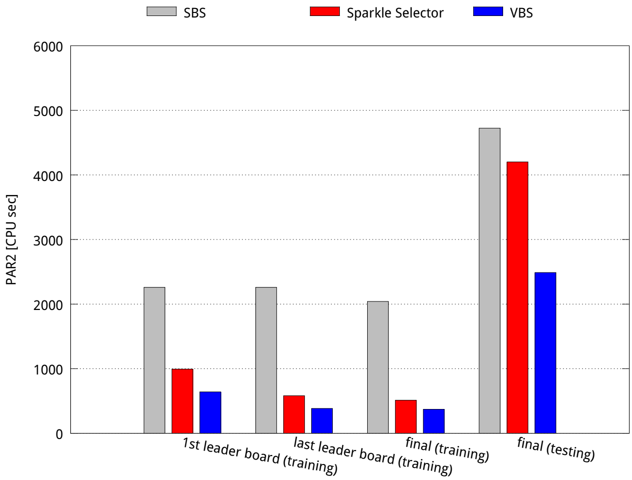
- SBS: 4740.02
- VBS: 2710.91
- Sparkle Selector: 4375.42

Official results:
Ranking according to marginal contribution on testing set, without hors-concours solvers

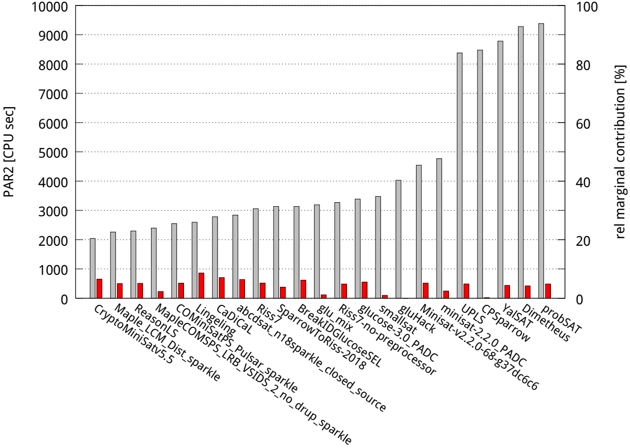
rank	solver (stand-alone rank)	rmc	amc
1	CryptoMiniSatv5.5 (1)	12.97%	0.0430
2	ReasonLS (2)	9.68%	0.0321
3	minisat-2.2.0_PADC (11)	9.07%	0.0301
4	glucose-3.0_PADC (14)	8.41%	0.0279
5	UPLS (16)	8.18%	0.0271
6	Riss7 (4)	7.81%	0.0259
7	probSAT (19)	6.99%	0.0232
8	CaDiCaL (6)	6.93%	0.0230
9	glu_mix (9)	6.20%	0.0205
10	BreakIDGlucoseSEL (10)	5.42%	0.0180

Additional insights:
Results with *hors-concours* solvers

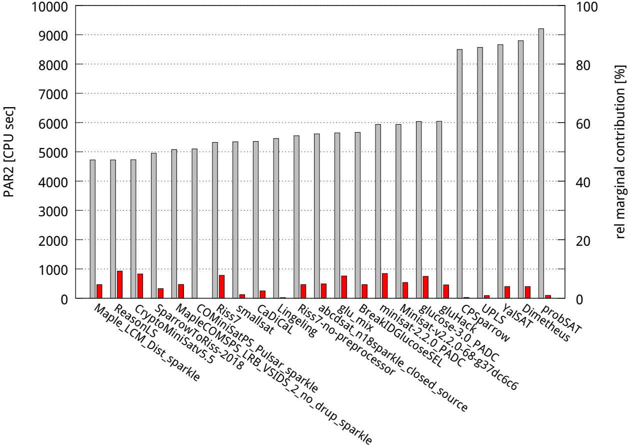
Improvement over time including hors-concours solvers



Stand-alone and relative marginal contribution on training set, with hors-concours solvers



Stand-alone and relative marginal contribution on testing set, with hors-concours solvers



Final results including hors-concours solvers, on testing set

PAR2 for SBS, VBS and Sparkle Selector

- SBS: 4724.03
- VBS: 2489.42
- Sparkle Selector: 4201.26

Ranking according to marginal contribution on testing set, with hors-concours solvers

Marginal contribution ranking (Top 10 solvers)

rank	solver (rank without hors-concourse solvers)	rmc	amc
1	ReasonLS (2)	9.30%	0.0533
2	minisat-2.2.0_PADC (3)	8.46%	0.0485
3	CryptoMiniSatv5.5 (1)	8.31%	0.0476
4	Riss7 (6)	7.83%	0.0448
5	glu_mix (9)	7.62%	0.0436
6	glucose-3.0_PADC (4)	7.49%	0.0429
7	Minisat-v2.2.0-68-g37dc6c6 (17)	5.40%	0.0309
8	*abcdsat_n18sparkle_closed_source (hors concours)*	*4.91%*	*0.0281*
9	*MapleCOMSPS_LRB_VSIDS_2_no_drup_sparkle (hors concours)*	*4.73%*	*0.0271*
10	Riss7-no-preprocessor (12)	4.67%	0.0268

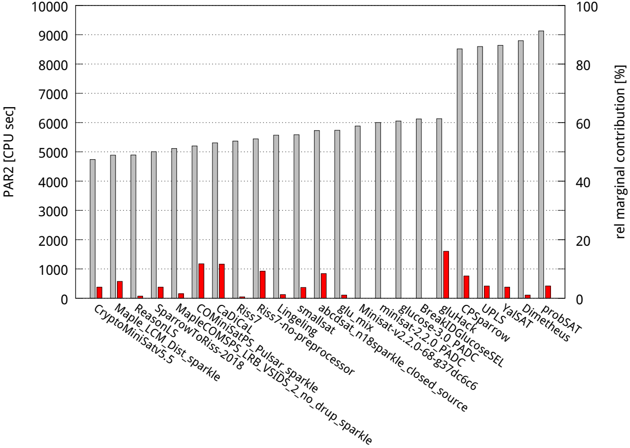
Additional insights: Training & testing on similar instances

What if we could train on instances from the same families as testing instances?

- testing set: 400 instances, 23 instance families
- split evaluation set from main track of 2018 SAT Competition into disjoint training and testing sets
  - for each instance family, ≈50% of instances → training set, remaining instances → testing set
- new training set: 195 instances from 23 families
- new testing set: 205 instances from 23 families

- PAR2 on new testing set:
  - SBS: 4739.87
  - VBS: 2498.68
  - Sparkle Selector: 3317.72 (75.3% of gap closed)

Stand-alone and relative marginal contribution on new testing set, with hors-concours solvers



Advantages of Sparkle challenge over traditional competition:

- makes it easier to gain recognition for specialised techniques
- better reflects and makes accessible state of the art
- provides incentive to improve true state of the art

Further use of Sparkle:

- Sparkle Planning Challenge 2019:
  http://ada.liacs.nl/events/sparkle-planning-19
- continuous solver evaluation (as community service)
- experimentation platform for algorithm selection, configuration, programming by optimisation (PbO)