Sparkle Planning Challenge 2019

Chuan Luo, Mauro Vallati and Holger H. Hoos

Universiteit Leiden
The Netherlands
&
University of Huddersfield
United Kingdom

ICAPS 2019, Berkeley, USA

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 / techniques
- ... has been substantially advanced by machine learning

Competitions ...

- ... have helped advance the state of the art in many fields (AI planning, SAT, ASP, machine learning, ...)

Competitions ...

- ... have helped advance the state of the art in many fields (AI planning, SAT, ASP, machine learning, ...)
- ... are mostly focused on single solvers, broad-spectrum performance
- ... often don't help to gain insights on state of the art, which is complex and variegated
- ... may not provide effective incentive to improve state of the art

A different kind of competition:

- solvers submitted to competition platform
- robust and effective per-instance selector built based on all solvers

A different kind of competition:

- solvers submitted to competition platform
- robust and effective per-instance selector built based on all solvers
- solver contributions to overall performance assessed based on (relative) marginal contribution
  (Xu, Hutter, HH, Leyton-Brown 2012; Luo, Vallati & Hoos – this event)
- full credit for contributions to selector performance goes to component solver authors

$\rightsquigarrow$ Sparkle Planning Challenge 2019 (Luo, Vallati & Hoos 2019 – this event)
$\rightsquigarrow$ Sparkle SAT Challenge 2018 (Luo & Hoos 2018)


XXXI Olimpíada Rio 2016


XXXI Olimpíada Rio 2016

Sparkle Planning Challenge 2019

- launched June 2018, leader board phase 18 March–12 April 2019, final results now!
- Settings as for IPC Agile track: 300 CPU-time seconds to solve, 8 GB of RAM.
- website: http://ada.liacs.nl/events/sparkle-planning-19

Planners submitted

- Aquaplanning; T. Balyo, D. Schreiber, P. Hegemann, J. Trautmann
- Cerberus; M. Katz
- dual-bfws; N. Lipovetzky, M. Ramirez, G. Frances, H. Geffner, C. Muise
- IPALAMA; D. Gnad, A. Torralba, M. Dominguez, C. Areces, F. Bustos
- Kronk; J. Seipp
- Madagascar; J. Rintanen
- MRW-RPG; R. Kuroiwa
- PASAR; N. Froleyks, T. Balyo, D. Schreiber
- PROBE; N. Lipovetzky, M. Ramirez, G. Frances, H. Geffner, C. Muise
- SYSU-Planner; Q. Yang, J. He, H.H. Zhuo

Testing domains

- Agricola IPC 2018
- **Baxter** A. Capitanelli, F. Mastrogiovanni, M. Maratea, M. Vallati
- CaveDiving IPC 2014
- **ChairGame** M. Vallati
- CityCar IPC 2014
- **Pipegrid** D. Schreiber
- Parking IPC 2008
- **UTC-distribution** L. Chrpa and M. Vallati
- Termes IPC 2018
- **Pizza** T. de la Rosa and R. Fuentetaja

Constructing the per-instance selector

- training set: 916 instances from 52 benchmark sets (domains), from deterministic tracks of 2014 and 2018 IPCs, and from testing domains
- split training set into *core training set* and *validating set*
- testing set: 100 instances from 10 domains
- no overlap in instances between training and testing sets

Constructing the per-instance selector

- training set: 916 instances from 52 benchmark sets (domains), from deterministic tracks of 2014 and 2018 IPCs, and from testing domains
- split training set into *core training set* and *validating set*
- testing set: 100 instances from 10 domains
- no overlap in instances between training and testing sets

- run AutoFolio (Lindauer *et al.* 2015) 100 times to obtain 100 per-instance selectors
  - train on core training set
  - choose selector with smallest PAR10 score on validating set

$\rightsquigarrow$ cutting-edge, robust algorithm selector construction in Sparkle

Assessing planner contributions

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

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

$$amc(s, I) = \begin{cases} log_{10} \frac{PAR10(P \setminus \{s\}, I)}{PAR10(P, I)} & PAR10(P \setminus \{s\}, I) > PAR10(P, I) \\ 0 & \text{else} \end{cases}$$

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

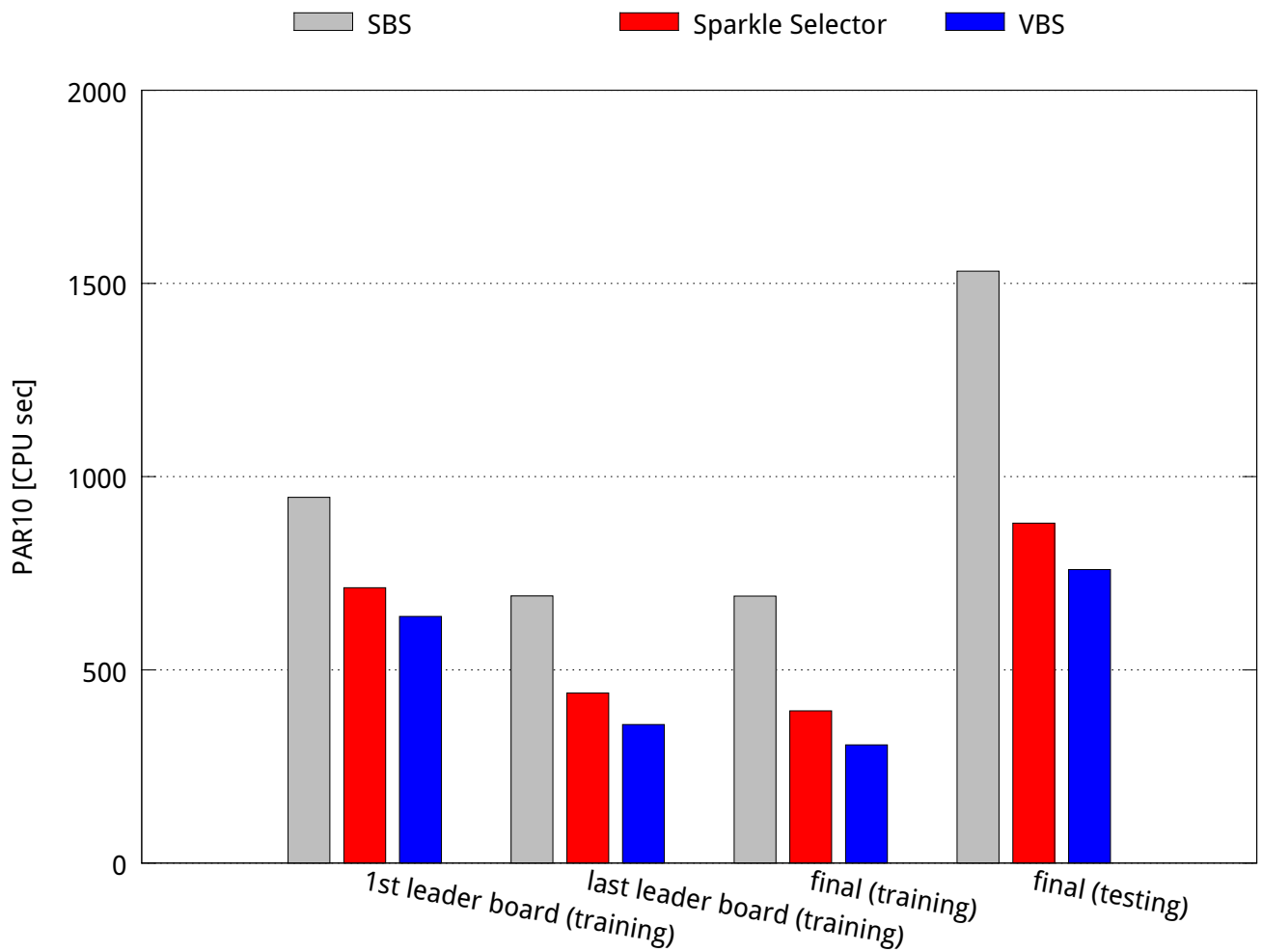
$$rmc(s, I) = \frac{amc(s)}{\sum_{s' \in S} amc(s')}$$

Final results on testing set

PAR10 in CPU sec
SBS, VBS and Sparkle Selector

- SBS: 1531.9 CPU sec
- VBS: 759.5 CPU sec
- Sparkle Selector: 879.7 CPU sec

Improvement over time



Official results:
Ranking according to marginal contribution on testing set

Official results:
Ranking according to marginal contribution on testing set

rank	solver (IPC rank)	rmc	amc

Official results: Ranking according to marginal contribution on testing set

rank	solver (IPC rank)	rmc	amc
1	PROBE (4)	34.77%	0.1401

Official results: Ranking according to marginal contribution on testing set

rank	solver (IPC rank)	rmc	amc
1	PROBE (4)	34.77%	0.1401
2	dual-bfws (1)	23.25%	0.0937

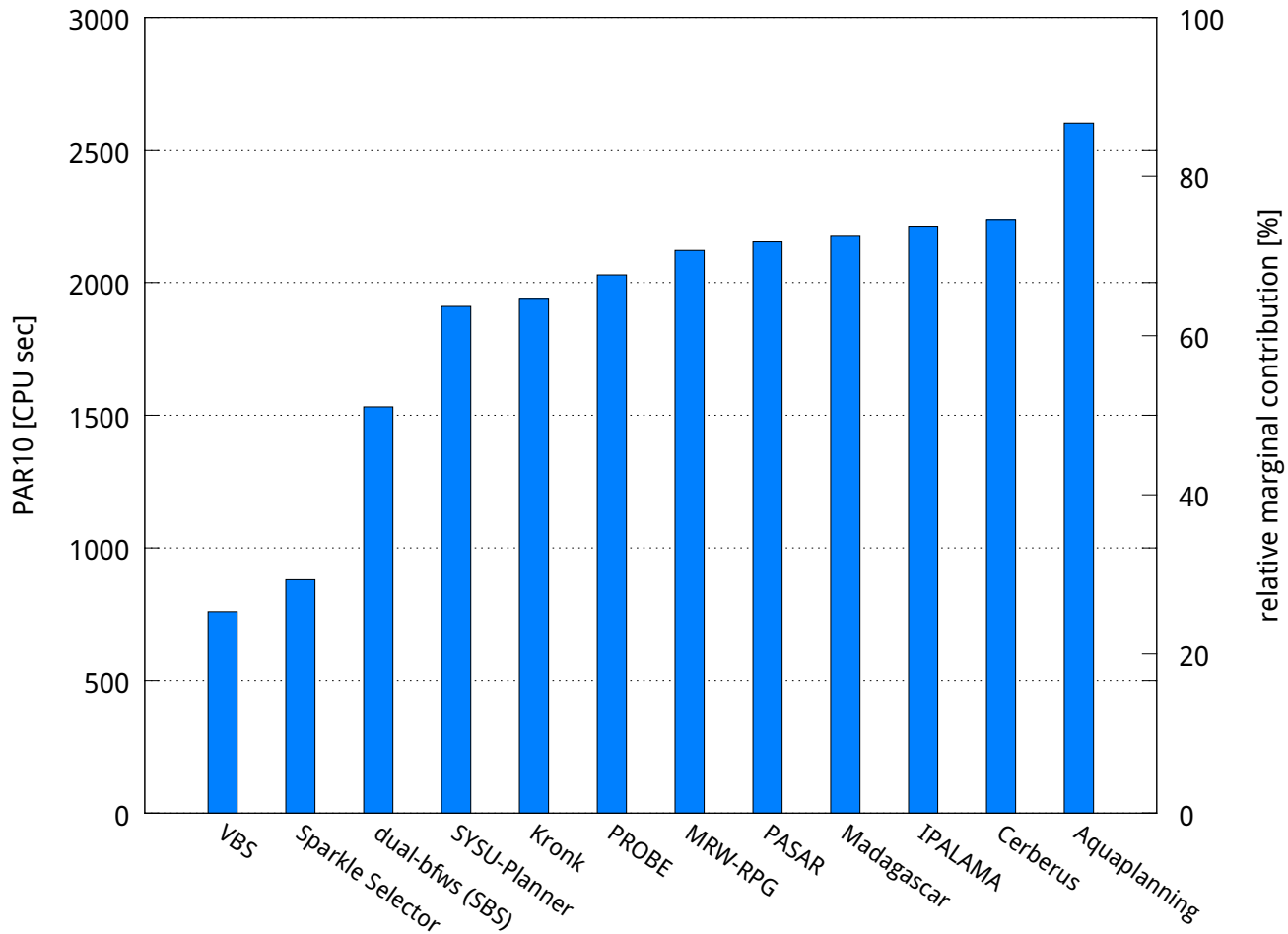
Official results: Ranking according to marginal contribution on testing set

rank	solver (IPC rank)	rmc	amc
1	PROBE (4)	34.77%	0.1401
2	dual-bfws (1)	23.25%	0.0937
3	PASAR (6)	22.13%	0.0892

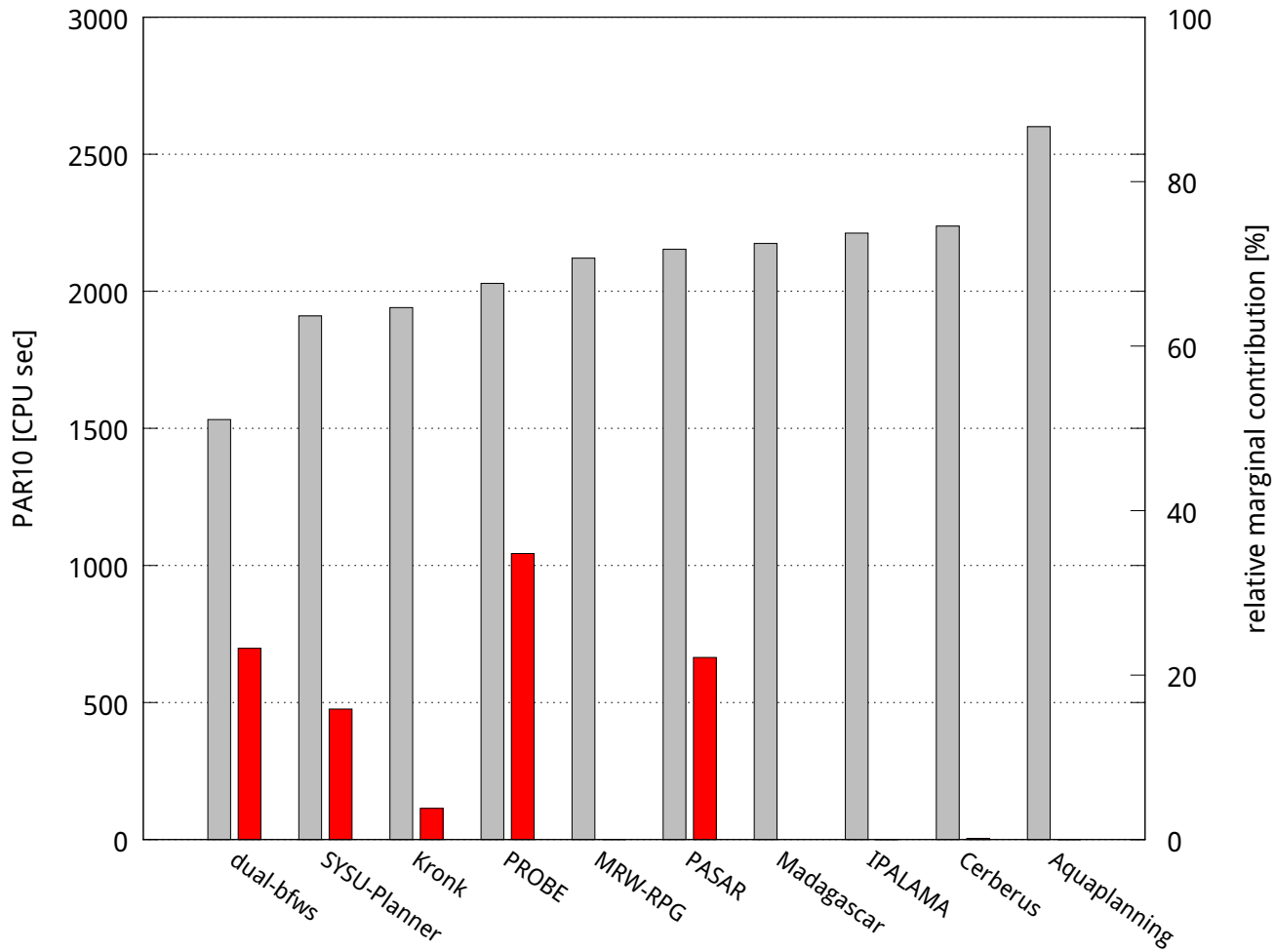
Official results: Ranking according to marginal contribution on testing set

rank	solver (IPC rank)	rmc	amc
1	PROBE (4)	34.77%	0.1401
2	dual-bfws (1)	23.25%	0.0937
3	PASAR (6)	22.13%	0.0892
4	SYSU-Planner (2)	15.86%	0.0639
5	Kronk (3)	3.80%	0.0153
6	Cerberus (9)	0.14%	0.0005
7	MRW-RPG (5)	0.01%	0.0001
8	IPALAMA (8)	0.01%	0.0001
9	Aquaplanning (10)	0.01%	0.0001
10	Madagascar (7)	0.01%	0.0001

Stand-alone and relative marginal contribution on testing set



Stand-alone and relative marginal contribution on testing set



Advantages of Sparkle challenge over traditional competition:

- can make it easier to gain recognition for specialised techniques
- can provide a better picture of the state of the art
- provides incentive to design innovative techniques

Advantages of Sparkle challenge over traditional competition:

- can make it easier to gain recognition for specialised techniques
- can provide a better picture of the state of the art
- provides incentive to design innovative techniques

Note:

- benchmark instances are getting more and more (structurally) different and complex
  $\rightsquigarrow$ Sparkle even more effective
- Detailed results:
  http://ada.liacs.nl/events/sparkle-planning-19