Open Algorithm Selection Challenge
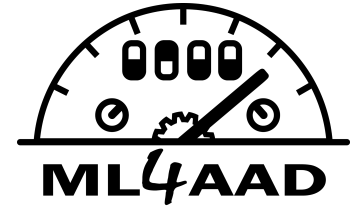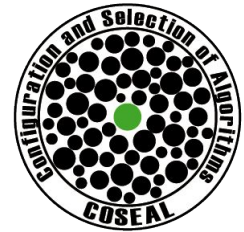
Lars Kotthoff Marius Lindauer Jan N. v. Rijn
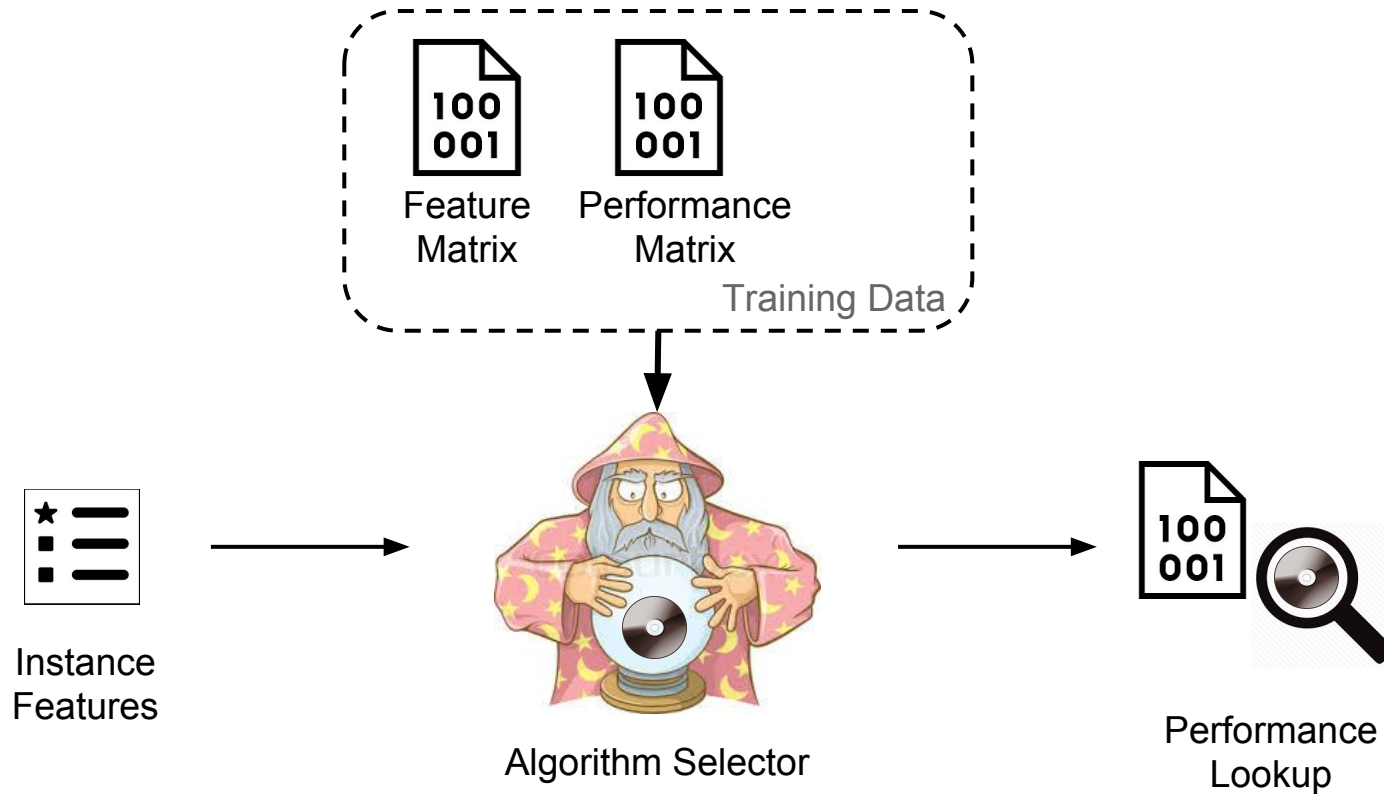
COSEAL Workshop 2017, Brussels

Algorithm Selection [Rice '76]

Task/
Instance

[12,
64,
49,
0,
3,
42,
89,
…]

Feature Construction

Predicted Algorithm

Solve Task

ASlib [Bischl et al. '16]

Feature Matrix

Performance Matrix

Training Data

Instance Features

Algorithm Selector

Performance Lookup

History: ICON Challenge on Algorithm Selection '15

- Submissions: AS systems

- But:

    - Time limit of 12h to train selector

    - Scenarios were known beforehand, but unknown train/test split

    - Only single pre-solver allowed

    - Dynamic selection of feature groups allowed

Open Algorithm Selection Challenge (OASC)

Differences to ICON AS Challenge (2015)

1. Submission: Predictions on unknown test data (2:1 training:test)

    a. No resource limits for predictions

    b. Closed-source submissions possible

2. Scenario source hidden

    a. Obfuscated scenario data

3. Scenarios with "solution quality" as objective

4. Flexible algorithm and feature schedules

OASC in a Nutshell

- 11 scenarios in ASlib format
  - 8 running time scenarios (6 new)
  - 3 solution quality scenarios (2 new)
- 8 submissions from 4 teams
  - All open-source
- Performance metric: *closed gap* (PAR10)
  - 1.0 : perfect algorithm selector (aka oracle or VBS)
 - 0.0 : best algorithm on average (no algorithm selection)

Submitted Systems

- ***zilla** [Xu et al. '08]
  - Pre-solving schedule + selector
  - "Dyn. sched." with per-instance dynamic schedules (only for running time scenarios)
- **ASAP** [Gonard et al. '16]
  - Pre-solving schedule + Selector
    - Joint optimization of both modules
  - "V2": fixed #pre-solvers, "V3": optimized #pre-solvers
- **Sunny** [Amadini et al. '15]
  - Per-instance algorithm scheduling strategy
  - "autok": optimized neighborhood size; "fkvar": + greedy forward feature selection
- **as-asl and as-rf** [Malone et al. '17; to appear in MLJ]
  - Stacked algorithm selector (EPM + classifier on top)
  - "asl": auto-sklearn; "rf": default random forest

Bado: BNSL-2016

- Structure learning in Bayesian networks [Melone et al. MLJ; to appear]

#Instances	1179
#Algorithms	8
#Features (#groups)	86 (7)
Objective	Time
VBS	219.86
Single Best	9017.07

Sunny-fkvar (Gap: 0.84)

AS-RF (Gap: 0.83)

ASAP.v3 (Gap: 0.80)

2   1   3
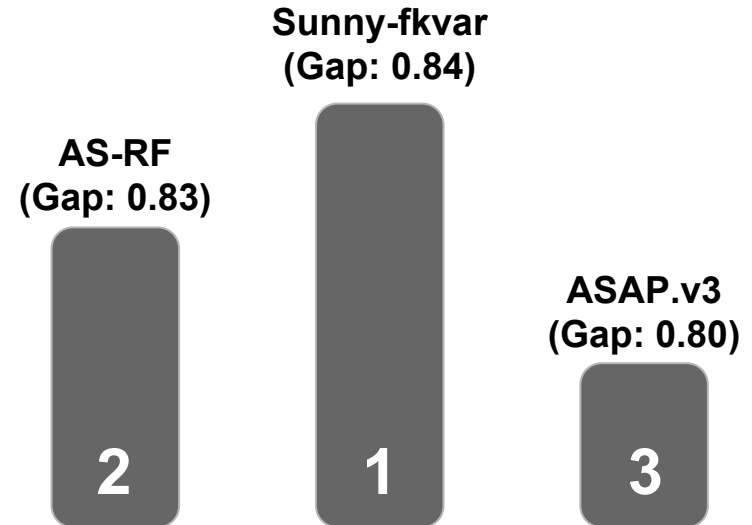
Camilla: CSP-Minizinc-Obj-2016

- Data from Minizinc CSP competition 2016

#Instances	**100**
#Algorithms	8
#Features (#groups)	95 (1)
Objective	**Solution Quality**
VBS	0.16
Single Best	0.28

ASAP.v3/2
(Gap: 0.98)

Sunny-fkvar
(Gap: 0.11)

1+2

3

Caren: CSP-Minizinc-Time-2016

- Data from Minizinc CSP competition 2016

#Instances	**100**
#Algorithms	8
#Features (#groups)	95 (1)
Objective	**Time**
VBS	26.27
Single Best	1605.36

**Sunny-fkvar
(Gap: 0.91)**

**Sunny-autok
(Gap: 0.64)**

***zilla (dyn. sched.)
(Gap: 0.63)**

2   1   3
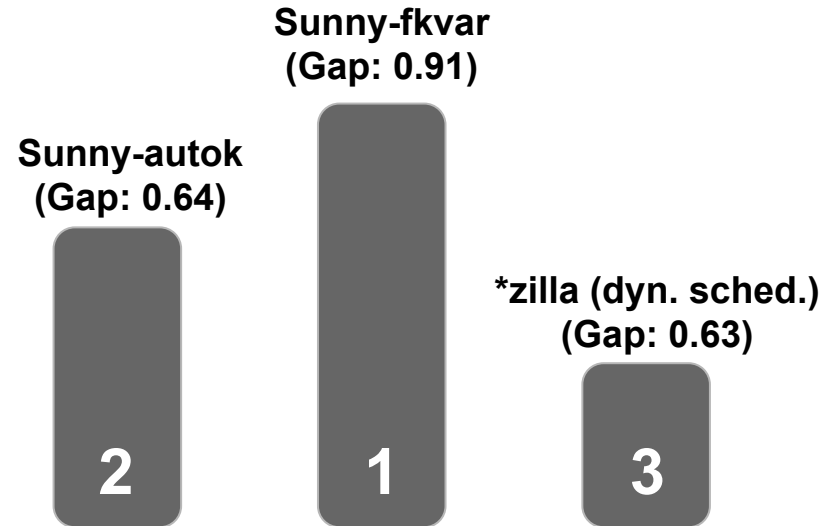
Magnus: MAXSAT-PMS-2016

- Data from MaxSAT Competition 2016
- Partial MaxSAT (PMS) industrial

#Instances	601
#Algorithms	19
#Features (#groups)	37 (1)
Objective	Time
VBS	38.31
Single Best	955.61

*zilla (dyn. sched.)
(Gap: 0.58)

*zilla
(Gap: 0.57)

Sunny-fkvar
(Gap: 0.57)

2　　**1**　　**3**
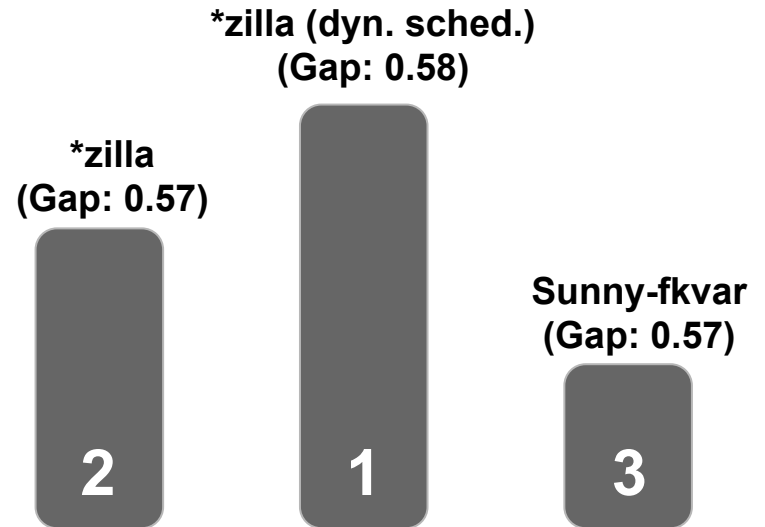
Monty: MAXSAT-WPMS-2016

- Data from MaxSAT Competition 2016
- **Weighted** Partial MaxSAT (PMS) industrial

#Instances	630
#Algorithms	18
#Features (#groups)	37 (1)
Objective	Time
VBS	101.34
Single Best	1572.00

Sunny-fkvar (Gap: 0.91)

ASAP.v2 (Gap: 0.83)

ASAP.v3 (Gap: 0.76)

2   1   3
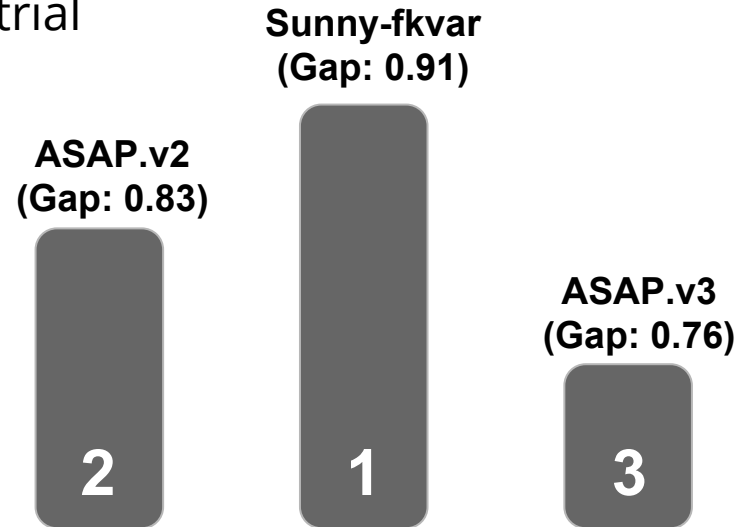
Mira: MIP-2016

- MIP solver evaluation '16 on MIPLIB2010

#Instances	218
#Algorithms	5
#Features (#groups)	143 (1)
Objective	Time
VBS	281.51
Single Best	3007.92

**ASAP.v2
(Gap: 0.50)**

**ASAP.v3
(Gap: 0.51)**

**AS-RF
(Gap: 0.49)**

2

1

3
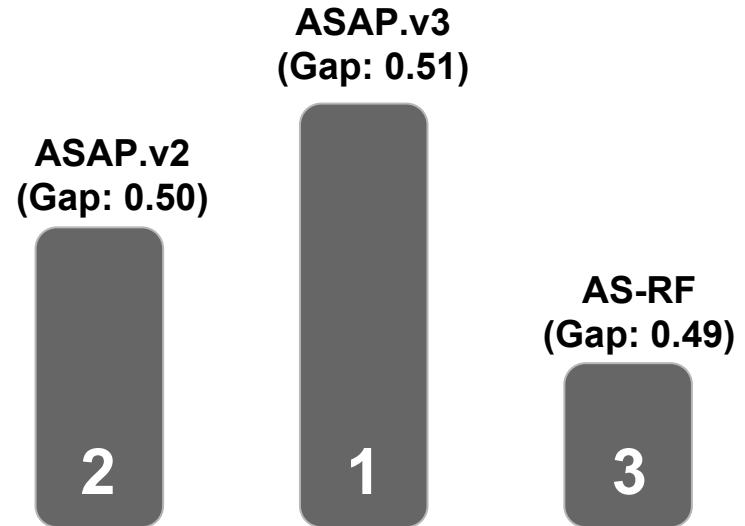
Oberon: OPENML-WEKA-2017

- Classifiers on data sets from OpenML

#Instances	**105**
#Algorithms	**19**
#Features (#groups)	**103** (1)
Objective	Accuracy
VBS	0.875
Single Best	0.855

**Sunny-fkvar
(Gap: -0.16)**

**Sunny-autok
(Gap: -0.30)**

**ASAP.v2/3
(Gap: -0.40)**

2

1

3
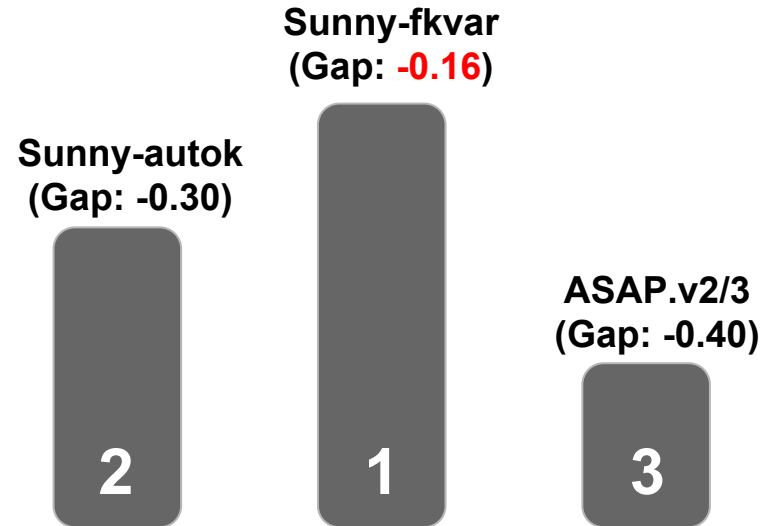
Quill: QBF-2016

- Performance data from QBF Evaluation 2016
  - Prenex CNF Track

#Instances	825
#Algorithms	24
#Features (#groups)	46 (1)
Objective	Running time
VBS	**9.99**
Single Best	2642.89

ASAP.v2
(Gap: 0.70)

Sunny-autok
(Gap: 0.84)

ASAP.v3
(Gap: 0.57)

2

1

3
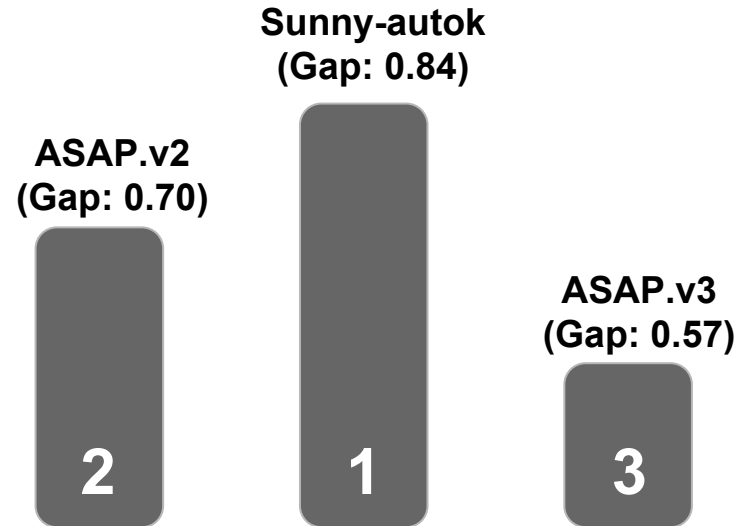
Svea: SAT12-ALL

- SAT instances (industrial + crafted + random)
- Training data for SATzilla 2012

#Instances	1614
#Algorithms	31
#Features (#groups)	**115** (10)
Objective	Running time
VBS	97.84
Single Best	2962.29

ASAP.v3
(Gap: 0.69)

ASAP.v2
(Gap: 0.68)

Sunny-fkvar
(Gap: 0.66)

2 1 3
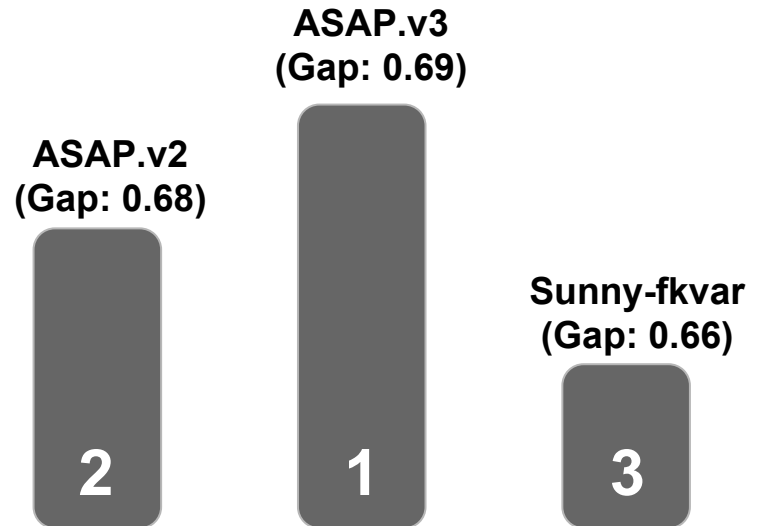
Sora: SAT03-16_INDU

- Industrial SAT instances
- Algorithms from 2016 SAT Competition

#Instances	2000
#Algorithms	10
#Features (#groups)	**483** (16)
Objective	Time
VBS	308.18
Single Best	3930.48

ASAP.v2
(Gap: 0.21)

*zilla [dyn. Sched.]
(Gap: 0.1706)

2+3

1
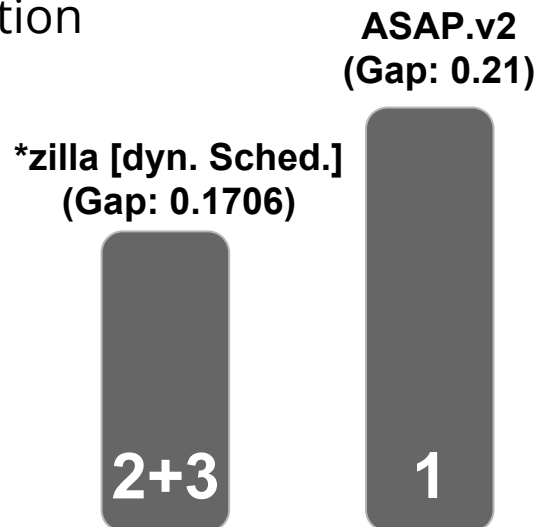
Titus: TTP-2016

- Travelling thief problem
- [Wagner et al. 2016]

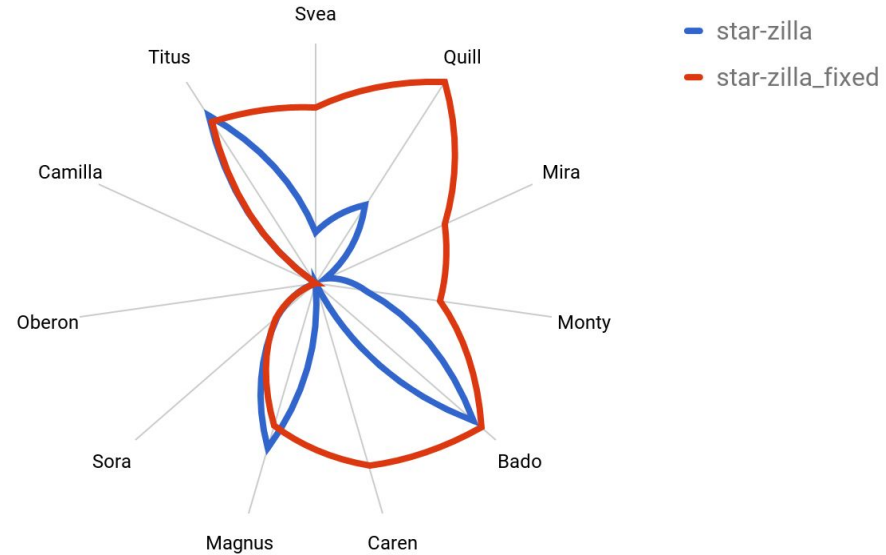| #Instances	**9720**
#Algorithms	22
#Features (#groups)	**50** (1)
Objective	Solution Quality
VBS	1.00
Single Best	0.96

**ASAP.v2/3
(Gap: 0.85)**

1+2

**Sunny-autok
(Gap: 0.81)**

3

Comment on *zilla

- Bug in official submission
- Resubmission after deadline
  - **Out of competition**
- Much better results
  - In particular on running time scenarios
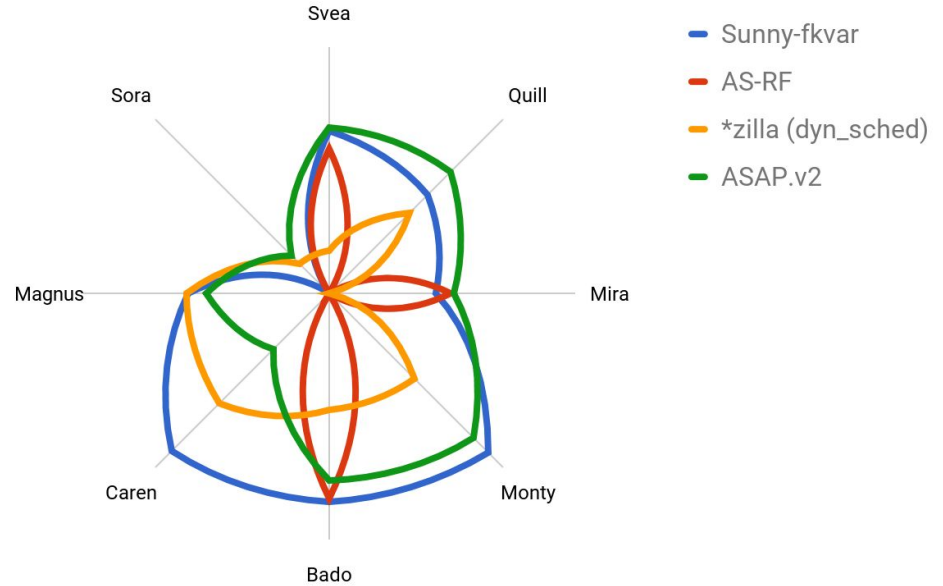 - Overall ranking does not change much

Overall Winners on Running Time

Overall Winners on Solution Quality

**ASAP.v2/3
(Gap: 0.47)**

**Sunny-fkvar
(Gap: 0.25)**

1+2

3

Titus

Oberon

Camilla
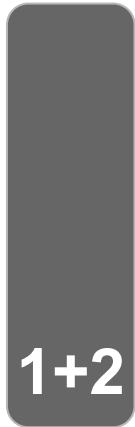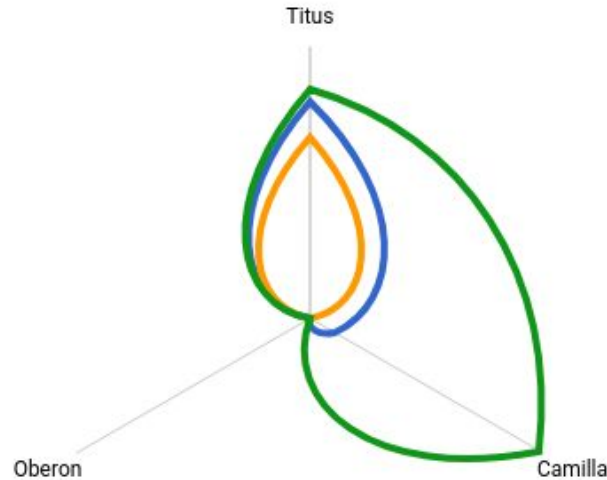
- Sunny-fkvar
- AS-RF
- *zilla (dyn_sched)
- ASAP.v2

Overall Winners

ASAP.v2
(Gap: 0.54)

ASAP.v3
(Gap: 0.5136)

Sunny-fkvar
(Gap: 0.5133)

2

1

3

- Sunny-fkvar
- AS-RF
- *zilla (dyn_sched)
- ASAP.v2

Bado, Camilla, Caren, Magnus, Mira, Monty, Oberon, Quill, Sora, Svea, Titus
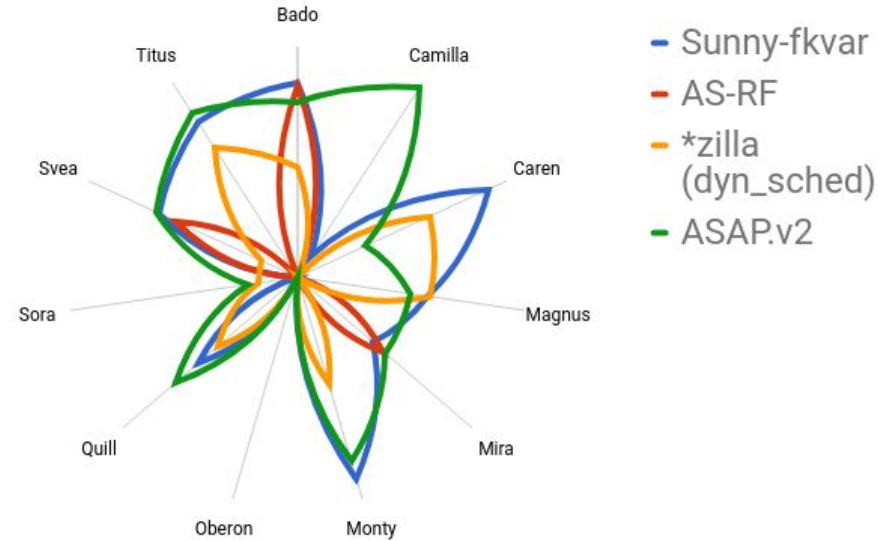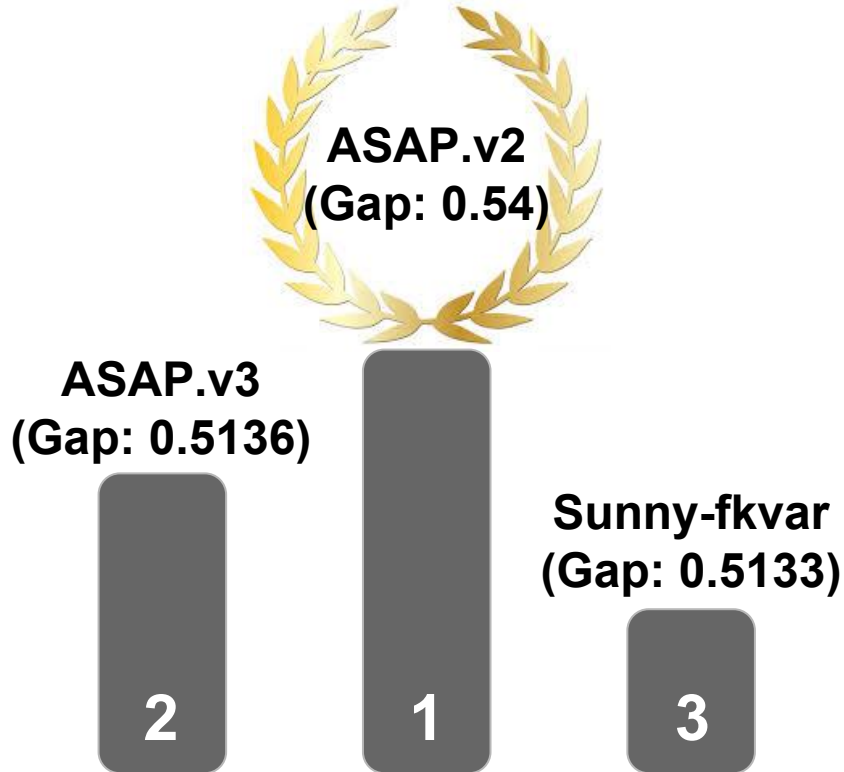
Take Home Messages

- Schedules are famous these days

- Averaged closed gap "only": 54%

  - Still hard on some scenarios

    - many features, small training set, …

  - Algorithm selection still not solved?

- 31.8 fold speedup on running time scenarios possible (SBS/VBS)

All detailed results and ASlib scenarios soon available at: **www.aslib.net**