MaxSAT Evaluation 2022

Fahiem Bacchus
University of Toronto

Matti Järvisalo
University of Helsinki

Jeremias Berg
University of Helsinki

Ruben Martins
Carnegie Mellon University

Andreas Niskanen
University of Helsinki

https://maxsat-evaluations.github.io/

SAT 2022, August 5, 2022

What is Maximum Satisfiability?

- Maximum Satisfiability (MaxSAT):
  - Clauses in the formula are either **soft** or **hard**
  - Hard clauses: **must** be satisfied
  - Soft clauses: **desirable** to be satisfied
  - Soft clauses may have weights

- **Goal**: Maximize (minimize) the sum of the weights of satisfied (unsatisfied) soft clauses
  - Equivalent to minimizing a linear objective function subject to a set of hard clauses.

Setup

Same structure as the one used in MaxSAT Evaluations 2017-2021:

- Source disclosure requirement:
  - Increase the dissemination of solver development
- Solver description using IEEE Proceedings style:
  - Better understanding of the techniques used by each solver
- Benchmark description using IEEE Proceedings style
  - Better understanding of the nature of each benchmark
- **Descriptions collected in proceeding, published at MSE website.**

News for this year

- Changes in the WCNF input format:
  - Removal of the p-line
  - Hard clauses marked with “h”:
    - Instead of sum of the weights of soft clauses + 1
- New special track:
  - “Incremental Track” on Incremental MaxSAT

News for this year

- Changes in the WCNF input format:
  - Removal of the p-line
  - Hard clauses marked with “h”:
    - Instead of sum of the weights of soft clauses + 1
- New special track:
  - “Incremental Track” on incremental MaxSAT

Evaluation tracks

Evaluation tracks:

- Complete:
  - Weighted
  - Unweighted
- No distinction between industrial and crafted benchmarks
- Incomplete:
  - Weighted
  - Unweighted
- **(New)** Incremental

Evaluation tracks

Evaluation tracks:

- Complete:
  - Weighted
  - Unweighted
- No distinction between industrial and crafted benchmarks
- Incomplete:
  - Weighted
  - Unweighted
- **(New)** Incremental

Evaluation tracks

Evaluation tracks:

- Complete:
  - Weighted
  - Unweighted
- No distinction between industrial and crafted benchmarks
- Incomplete:
  - Weighted
  - Unweighted
- **(New)** Incremental

Execution environment

MSE 2022 was run on the StarExec cluster:

- `https://www.starexec.org/`
- Intel(R) Xeon(R) CPU E5-2609 0 @ 2.40GHz
- 10240 KB Cache, 128 GB Memory
- Two solvers per node

Execution environment:

- Complete:
  - Time limit: 3600 seconds
  - Memory limit: 32 GB
- Incomplete track:
  - Two time limits: 60 seconds and 300 seconds
  - Memory limit: 32 GB

Execution environment

Incremental track run on a computing cluster at the University of Helsinki, part of the Finnish Computing Competence Infrastructure (FCCI).

- `https://www2.helsinki.fi/en/infrastructures/fcci`
- Intel(R) Xeon(R) CPU E5-2670 0 @ 2.60GHz
- 20 MB Cache, 57 GB Memory

Execution environment:

- Time limit: 7200 seconds
- Memory limit: 32 GB

Benchmark Selection

- Complete track:
  - As described earlier.

- Incomplete track:
  - Hard instances of the complete track.
    - Instances that cannot be solved optimally in 300 seconds by any participants of the complete tracks.

- Incremental track:
  - For each benchmark application, exclude instances which are easy for alternative approaches: solved in < 1 second or via one MaxSAT call.
  - Sample 100 instances from this set of instances.

MSE 2022 benchmarks

Complete track:

- Unweighted (594 instances)
  - 46 different domains
- Weighted (607 instances)
  - 46 different domains

Incomplete track:

- Unweighted (179 instances)
- Weighted (197 instances)

Incremental track:

- Five benchmark applications, 100 instances per application

Complete Track

Solvers

MaxSAT approaches in MSE 2022:

Solver	Hitting Set	Unsat-based	Sat-Unsat	Other
CASHWMaxSAT		✓		ILP
Exact		✓		PB
MaxCDCL				B&B
EvalMaxSAT		✓		
MaxHS	✓	✓	✓	ILP
UWrMaxSat		✓	✓	ILP
Open-WBO		✓		
CGSS		✓		

- Solvers are using a combination of techniques!
- ILP is becoming popular to solve some instances!

Solvers

New solvers:

- **MaxCDCL** by Jordi Coll et al.

  Chu-Min Li, Zhenxing Xu, Jordi Coll, Felip Manyá, Djamal Habet, Kun He: Combining Clause Learning and Branch and Bound for MaxSAT. CP 2021: 38:1-38:18

- **CGSS** by Hannes Ihalainen, Jeremias Berg, Matti Järvisalo

  Hannes Ihalainen, Jeremias Berg, Matti Järvisalo: Refined Core Relaxation for Core-Guided MaxSAT Solving. CP 2021: 28:1-28:19

Results

Unweighted: 594 instances

Solver	#Solved	Time (Avg)

Weighted: 607 instances

Solver	#Solved	Time (Avg)

Results

Unweighted: 594 instances

Solver	#Solved	Time (Avg)
CASHWMaxSAT-CorePlus	438	181.69
CASHWMaxSAT-Plus	433	184.96
UWrMaxSat-SCIP	432	308.53

Weighted: 607 instances

Solver	#Solved	Time (Avg)

Results

Unweighted: 594 instances

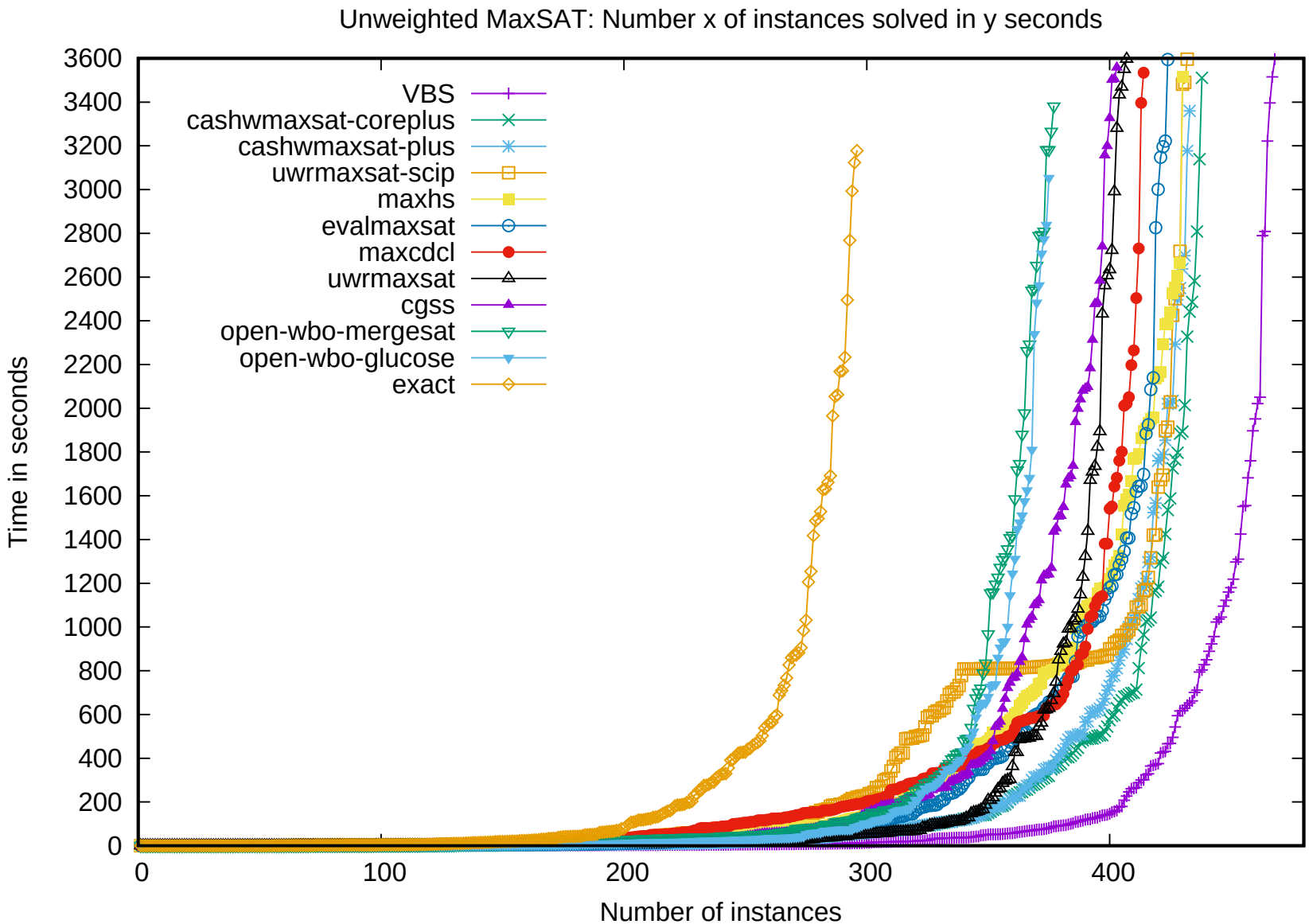
Solver	#Solved	Time (Avg)
CASHWMaxSAT-CorePlus	438	181.69
CASHWMaxSAT-Plus	433	184.96
UWrMaxSat-SCIP	432	308.53

Weighted: 607 instances

Solver	#Solved	Time (Avg)
CASHWMaxSAT-CorePlus	438	304.82
CASHWMaxSAT-Plus	435	303.16
UWrMaxSat-SCIP	427	305.70

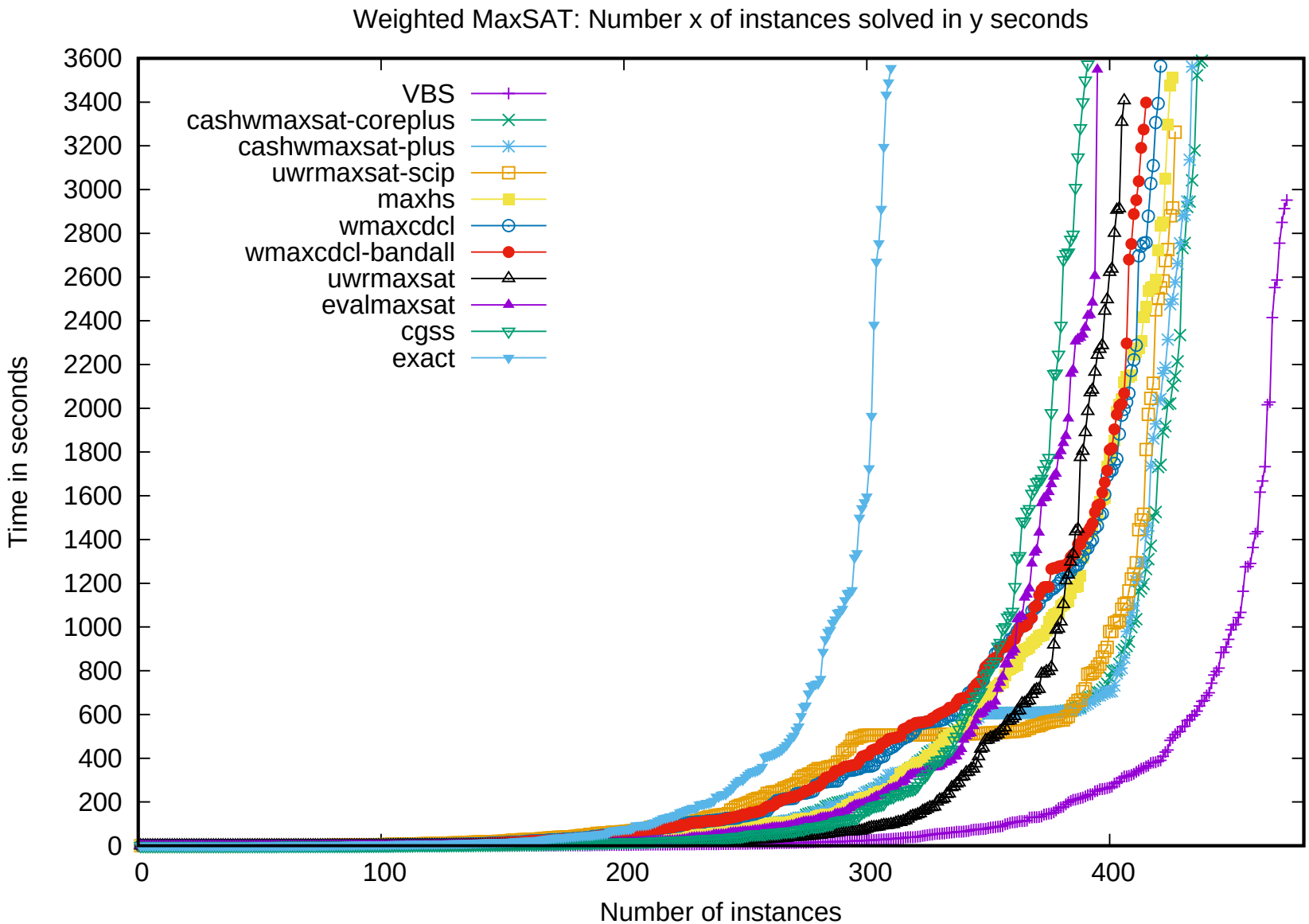
Detailed Results

Unweighted



Detailed Results

Weighted



How helpful are ILP solvers?

	Unweighted		Weighted	
Solver	800s	1 hour	500s	1 hour
UWrMaxSat				
UWrMaxSat+SCIP				

- SCIP + UWrMaxSat solves more instances than pure UWrMaxSat over 1 hour.
- UWrMaxSat solves more than pure SCIP over shorter period.
- ILP solvers can improve the performance of MaxSAT solvers!

How helpful are ILP solvers?

	Unweighted		Weighted	
Solver	800s	1 hour	500s	1 hour
UWrMaxSat		407		407
UWrMaxSat+SCIP		432		427

- SCIP + UWrMaxSat solves more instances than pure UWrMaxSat over 1 hour.
- UWrMaxSat solves more than pure SCIP over shorter period.
- ILP solvers can improve the performance of MaxSAT solvers!

How helpful are ILP solvers?

	Unweighted		Weighted	
Solver	800s	1 hour	500s	1 hour
UWrMaxSat	378	407	352	407
UWrMaxSat+SCIP	339	432	297	427

- SCIP + UWrMaxSat solves more instances than pure UWrMaxSat over 1 hour.
- UWrMaxSat solves more than pure SCIP over shorter period.
- ILP solvers can improve the performance of MaxSAT solvers!

How helpful are ILP solvers?

	Unweighted		Weighted	
Solver	800s	1 hour	500s	1 hour
UWrMaxSat	378	407	352	407
UWrMaxSat+SCIP	339	432	297	427

- SCIP + UWrMaxSat solves more instances than pure UWrMaxSat over 1 hour.
- UWrMaxSat solves more than pure SCIP over shorter period.
- ILP solvers can improve the performance of MaxSAT solvers!

SCIP with other solvers

Results

Unweighted: SCIP (800s) + other solver (2800s)

Solver	#Solved
SCIP + MaxHS	430 → 446
SCIP + EvalMaxSAT	426 → 439
CASHWMaxSAT-CorePlus	438

Weighted: SCIP (500s) + other solver (3100s)

Solver	#Solved
SCIP + MaxHS	426 → 440
SCIP + WMaxCDCL	422 → 440
CASHWMaxSAT-CorePlus	438

Note that CASHWMaxSAT-CorePlus already uses SCIP.

Incomplete Track

Ranking for incomplete tracks

- Incomplete score: computed by the sum of the ratios between the best solution found by a given solver and the best known solution:
  - $\text{SCORE}(\textit{solver}, i) = \frac{(\text{cost of best known solution for i} + 1)}{(\text{cost of solution for i found by solver} + 1)}$
  - For an instance i score is 0 if no solution was found by that solver
  - For each instance the incomplete score is a value in $[0, 1]$
- Ranking based on average of all scores.

Solvers

- DT-HyWalk (new)
  - Local search and a SAT solver based algorithm (TT-Open-WBO-Inc)
- noSAT-MaxSAT (new)
  - Local search without a SAT solver.
- NuWLS-c (new)
  - Local search and a SAT solver based algorithm (TT-Open-WBO-Inc)
- Exact
- Loandra
- Open-WBO-inc (two variants)*
- TT-Open-WBO-Inc (three variants)

Only weighted track

Incomplete track: Unweighted (60 seconds)

Solver	Score (avg)
NuWLS-c	0.807
DT-Hywalk	0.773
TT-Open-WBO-inc (g variant)	0.769
TT-Open-WBO-inc (is variant)	0.764
SatLike-c (2021 variant)	0.764
TT-Open-WBO-inc (i variant)	0.756
Loandra	0.634
noSAT-MaxSAT	0.517
Exact	0.426

Best performing solver in 2021

Incomplete track: Unweighted (300 seconds)

Solver	Score (avg)
NuWLS-c	0.895
DT-Hywalk	0.874
TT-Open-WBO-inc (i variant)	0.868
TT-Open-WBO-inc (is variant)	0.866
TT-Open-WBO-inc (g variant)	0.864
SatLike-c (2021 variant)	0.863
Loandra	0.781
noSAT-MaxSAT	0.596
Exact	0.481

Best performing solver in 2021

Incomplete track: Weighted (60 seconds)

Solver	Score (avg)
NuWLS-c	0.759
SatLike-ck (2021 variant)	0.750
DT-Hywalk	0.732
TT-Open-WBO-inc (g variant)	0.728
TT-Open-WBO-inc (is variant)	0.726
TT-Open-WBO-inc (i variant)	0.720
Loandra	0.693
Open-WBO-inc-comp	0.689
Open-WBO-inc-satlike	0.654
Exact	0.475
noSAT-MaxSAT	0.442

Best performing solver in 2021

Incomplete track: Weighted (300 seconds)

Solver	Score (avg)
NuWLS-c	0.846
Loandra	0.814
Loandra (2021 variant)	0.806
DT-Hywalk	0.798
TT-Open-WBO-inc (i variant)	0.791
TT-Open-WBO-inc (g variant)	0.777
TT-Open-WBO-inc (is variant)	0.776
Open-WBO-inc-comp	0.774
Open-WBO-inc-satlike	0.750
Exact	0.541
noSAT-MaxSAT	0.512

Best performing solver in 2021

Observations

- Improvements to last year in both weighted and unweighted solvers.
- Hybrid solvers combining many different algorithms seem most effective.
- Size of benchmarks poses increasingly severe challenges.
  - For some solvers, even reading some benchmarks within 60s was challenging

Incremental Track

Incremental track: Overview

- Various problem domains call for procedures where a sequence of related optimization problems is solved
  - adding or removing constraints, modifying objective function
- Incremental solving: avoid computation from scratch, reuse information from previous calls
- MaxSAT solving techniques can be adapted to incremental settings
  - nontrivial from both theoretical and practical perspectives
- New experimental track: motivate and support the development of incremental MaxSAT solvers and applications
  - facilitated by IPAMIR: generic API for incremental MaxSAT solving, built on IPASIR (API for incremental SAT)

Incremental track: Submissions

- 5 benchmark submissions:
  - **Bi-objective Boolean optimization**: adding hard clauses
  - **MLIC-SeeSaw**: adding hard clauses + assumptions
  - **Extension enforcement in abstract argumentation**: adding hard clauses
  - **Learning boosted decision trees via AdaBoost**: changing weights of soft literals
  - **Proof obligations in bit-level PDR**: assumptions
- 3 solver submissions:
  - **EvalMaxSAT**: core-guided
  - **iMaxHS**: implicit hitting set based
  - **UWrMaxSat** (2 versions): core-guided (+ ILP)

Incremental track: Results

On each benchmark application, **rank** of solver determined by **number of solved instances** (out of 100). Ties are broken by cumulative CPU time.

Solver	rank (number of solved instances)				
	BiOptSat	SeeSaw	ExtEnf	AdaBoost	PDR
EvalMaxSAT	4 (28)	**1** (19)	2 (40)	4 (16)	**1** (44)
iMaxHS	3 (45)	2 (18)	**1** (48)	**1** (23)	3 (36)
UWrMaxSat	**1** (50)	3 (6)	3 (38)	2 (17)	2 (38)
UWrMaxSat+SCIP	2 (50)	N/A	4 (37)	3 (17)	4 (31)

Incremental track: Results

On each benchmark application, **rank** of solver determined by **number of solved instances** (out of 100). Ties are broken by cumulative CPU time.

Solver	rank (number of solved instances)				
	BiOptSat	SeeSaw	ExtEnf	AdaBoost	PDR
EvalMaxSAT	4 (28)	**1** (19)	2 (40)	4 (16)	**1** (44)
iMaxHS	3 (45)	2 (18)	**1** (48)	**1** (23)	3 (36)
UWrMaxSat	**1** (50)	3 (6)	3 (38)	2 (17)	2 (38)
UWrMaxSat+SCIP	2 (50)	N/A	4 (37)	3 (17)	4 (31)

Incremental track: Observations

- Solver performance application-dependent
  - EvaIMaxSAT, iMaxHS, and UWrMaxSat ranked first on some benchmark, all solvers ranked second on some benchmark
- Some benchmarks extremely hard to solve: under a fifth of instances solved, no matter which solver chosen
- Non-adaptive benchmarks: the sequence of IPAMIR calls depends on the results of previous solve calls
  - How to clearly rank solvers in this case?
- Initially all solvers had bugs on some benchmarks: testing is tedious as no concrete applications were available
  - New benchmark applications will serve as a basis for testing solvers
  - Unit tests and fuzzers for IPAMIR?

Hoping for more incremental solvers and benchmark applications next year!

Webpages

MaxSAT Evaluation 2022 webpage

`https://maxsat-evaluations.github.io/2022/`

- Detailed results for each instance
- Description of the solvers and benchmarks
- Source code of the solvers
- Benchmarks and log files.

MaxSAT Lib

`http://www.cs.toronto.edu/maxsat-lib/`

- Collection of all MaxSat instances submitted to the Evaluation.
- Best known costs of benchmarks.

Thanks

Thanks to everyone that contributed solvers and benchmarks!
Without you this evaluation would not be possible!

Thanks to StarExec for allowing us to use their cluster:

https://www.starexec.org/

Thanks to FCCI for supporting the incremental track with computational and data storage resources:

https://www2.helsinki.fi/en/infrastructures/fcci