

MaxSAT Evaluation 2021

Fahiem Bacchus
University Toronto

Jeremias Berg
University Helsinki

Matti Järvisalo
University Helsinki

Ruben Martins
CMU

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

SAT 2021, July 8, 2021

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

Setup

Same structure as the one used in MaxSAT Evaluations 2017-2020:

- ▶ 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.

Setup

Same structure as the one used in MaxSAT Evaluations 2017-2020:

- ▶ 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.**

New features

- ▶ New benchmark selection
 - ▶ Filter benchmarks known to have optimum cost 0 and known random domains.
 - ▶ 48 previously unknown 0 cost instances in the evaluation sets.
 - ▶ For every other domain, randomly sample up to 12 ($600/\text{\#domains}$) benchmarks.
 - ▶ Prioritize benchmarks that have not appeared in MSE 18-20
 - ▶ Sample twice as many (24) from domains new for this year.
 - ▶ 73% of the benchmarks in the complete track have not been seen in the last three years!

New features

- ▶ New benchmark selection
 - ▶ Filter benchmarks known to have optimum cost 0 and known random domains.
 - ▶ 48 previously unknown 0 cost instances in the evaluation sets.
 - ▶ For every other domain, randomly sample up to 12 ($600/\text{\#domains}$) benchmarks.
 - ▶ Prioritize benchmarks that have not appeared in MSE 18-20
 - ▶ Sample twice as many (24) from domains new for this year.
- ▶ 73% of the benchmarks in the complete track have not been seen in the last three years!

New features

- ▶ New benchmark selection
 - ▶ Filter benchmarks known to have optimum cost 0 and known random domains.
 - ▶ 48 previously unknown 0 cost instances in the evaluation sets.
 - ▶ For every other domain, randomly sample up to 12 ($600/\text{\#domains}$) benchmarks.
 - ▶ Prioritize benchmarks that have not appeared in MSE 18-20
 - ▶ Sample twice as many (24) from domains new for this year.
 - ▶ 73% of the benchmarks in the complete track have not been seen in the last three years!

New(ish) features

v 1 -2 -3 4 -5 ... -4567 4568 ... \longrightarrow v 10010...01...

- ▶ Compact v-line:

- ▶ Print the model as a sequence of '0' and '1' characters instead of using variable numbers
- ▶ Significant reduction in the size of the logs

- ▶ Incomplete score:

- ▶ Consider the best known solution (instead of the best solution found by the incomplete solvers)
- ▶ Keep a database of best known solution and update it periodically

New(ish) features

v 1 -2 -3 4 -5 ... -4567 4568 ... \longrightarrow v 10010...01...

- ▶ Compact v-line:
 - ▶ Print the model as a sequence of '0' and '1' characters instead of using variable numbers
 - ▶ Significant reduction in the size of the logs

- ▶ Incomplete score:
 - ▶ Consider the best known solution (instead of the best solution found by the incomplete solvers)
 - ▶ Keep a database of best known solution and update it periodically
 - ▶ Can be found at MaxSAT Lib:
<http://www.cs.toronto.edu/maxsat-lib/>

Evaluation tracks

Evaluation tracks:

- ▶ Complete:
 - ▶ Weighted
 - ▶ Unweighted
- ▶ No distinction between industrial and crafted benchmarks
- ▶ Incomplete:
 - ▶ Weighted
 - ▶ Unweighted

Evaluation tracks

Evaluation tracks:

- ▶ Complete:
 - ▶ Weighted
 - ▶ Unweighted
- ▶ No distinction between industrial and crafted benchmarks
- ▶ Incomplete:
 - ▶ Weighted
 - ▶ Unweighted

Evaluation tracks

Evaluation tracks:

- ▶ Complete:
 - ▶ Weighted
 - ▶ Unweighted
- ▶ No distinction between industrial and crafted benchmarks
- ▶ Incomplete:
 - ▶ Weighted
 - ▶ Unweighted

Execution environment

MSE 2021 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

Benchmark Selection

- ▶ Complete track:
 - ▶ As described earlier.
 - ▶ weighted track: 623 benchmarks
 - ▶ unweighted track: 561 benchmarks
- ▶ Incomplete track:
 - ▶ Hard instances of the complete track.
 - ▶ weighted track: 151 instances
 - ▶ unweighted track: 155 benchmarks

Benchmark Selection

- ▶ Complete track:
 - ▶ As described earlier.
 - ▶ weighted track: 623 benchmarks
 - ▶ unweighted track: 561 benchmarks
- ▶ Incomplete track:
 - ▶ Hard instances of the complete track.
 - ▶ weighted track: 151 instances
 - ▶ unweighted track: 155 benchmarks

Benchmark Selection

- ▶ Complete track:
 - ▶ As described earlier.
 - ▶ weighted track: 623 benchmarks
 - ▶ unweighted track: 561 benchmarks
- ▶ Incomplete track:
 - ▶ Hard instances of the complete track.
 - ▶ Instances that cannot be solved optimally in 300 seconds by any participants of the complete tracks.
 - ▶ weighted track: 151 instances
 - ▶ unweighted track: 155 benchmarks

Benchmark Selection

- ▶ Complete track:
 - ▶ As described earlier.
 - ▶ weighted track: 623 benchmarks
 - ▶ unweighted track: 561 benchmarks
- ▶ Incomplete track:
 - ▶ Hard instances of the complete track.
 - ▶ Instances that cannot be solved optimally in 300 seconds by any participants of the complete tracks.
 - ▶ weighted track: 151 instances
 - ▶ unweighted track: 155 benchmarks

New benchmarks

- ▶ Planning with Learned Binarized Neural Networks
- ▶ University Course Timetabling
- ▶ Learning Optimal Decision Trees and Boosted Trees
- ▶ Functional Sequences Maximizing the Sustained Switching Activity in a Pipelined Processor

Thank you to everyone who submitted benchmarks!

New benchmarks

- ▶ MaxSAT is being used in many applications!
 - ▶ 4 new domains this year
- ▶ Benchmark size is getting large:
 - ▶ All benchmarks **22 GB** (after gzip compression)
 - ▶ A challenge especially for incomplete solvers.

Complete Track

Complete track: Unweighted

MaxSAT approaches in MSE 2021:

Solver	Hitting Set	Unsat-based	Sat-Unsat
EvalMaxSAT (2020)		✓	
MaxHS	✓		
Exact		✓(PB-based)	
CASHWMaxSAT		✓	
Pacose (2020)			✓
UWrMaxSAT		✓	✓
Open-WBO		✓	

- ▶ Diverse approaches in MaxSAT!
- ▶ Each approach is important and can solve different applications!

Complete track: Unweighted

New and/or improved solvers:

- ▶ **MaxHS** by Fahiem Bacchus, University of Toronto
Improvements in the detection of abstract cores. Uses Cadical as the underlying SAT solver.
- ▶ **UWrMaxSAT** by Marek Piotrów, University of Wroclaw
Extended with generalized boolean multilevel optimization and a better detection of reusable cardinality structures.
- ▶ **Exact** (new) by Jo Devriendt, KU Leuven
Uses cutting-planes learning (CDCPL) and an OLL-like procedure extended to PBO.
- ▶ **CASHWMaxSAT** (new) by Zhendong Lei et al.
Based on UWrMaxSAT. Solves small instances with an IP solver.

Complete track: Unweighted

561 instances

Solver	#Solved	Time (Avg)

Complete track: Unweighted

561 instances

Solver	#Solved	Time (Avg)
EvalMaxSAT	411	145.48

Complete track: Unweighted

561 instances

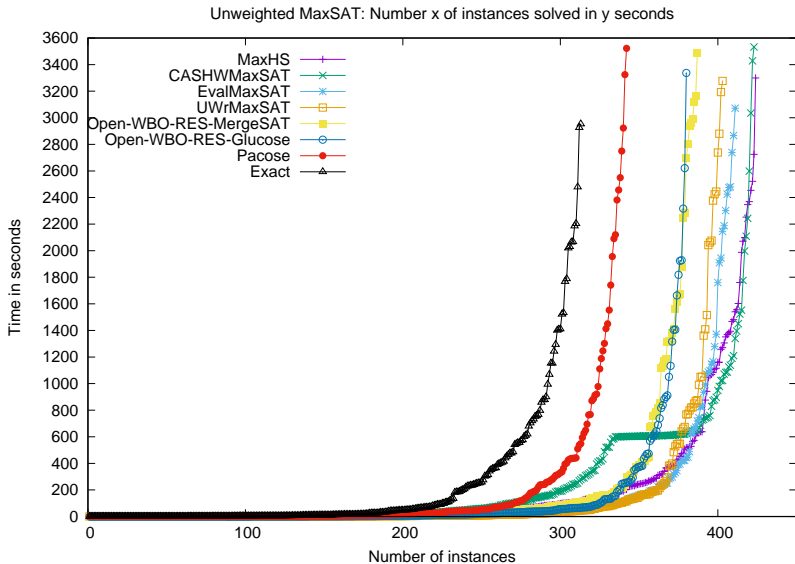
Solver	#Solved	Time (Avg)
CASHWMaxSAT	423	241.41
EvalMaxSAT	411	145.48

Complete track: Unweighted

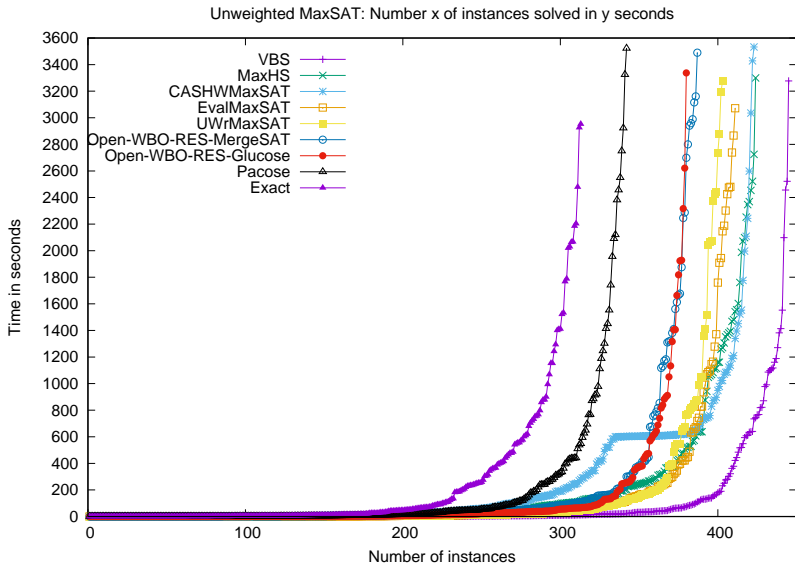
561 instances

Solver	#Solved	Time (Avg)
MaxHS	424	194.62
CASHWMaxSAT	423	241.41
EvalMaxSAT	411	145.48

Complete track: Unweighted



Complete track: Unweighted



Complete track: Unweighted

- ▶ Similar performance with respect to best solvers of latest years:
 - ▶ 2021: MaxHS solved 424 instances, 194.62 seconds (average)
 - ▶ 2020: MaxHS solved 425 instances, 192.35 seconds (average)
 - ▶ 2018: RC2 solved 399 instances, 169.99 seconds (average)
- ▶ VBS solves 445 instances:
 - ▶ VBS is getting closer to the best solver
 - ▶ Only 21 more instances!
- ▶ Closer look at the VBS:

UWrMaxSAT	98	MaxHS	60
EvalMaxSAT	98	CASHWMaxSAT	23
Open-WBO-RES-Glucose	88	Open-WBO-RES-MergeSAT	9
Pacose	62	Exact	7

Complete track: Weighted

MaxSAT approaches in MSE 2021:

Solver	Hitting Set	Unsat-based	Sat-Unsat
MaxHS	✓		
Exact		✓(PB-based)	
CASHWMaxSAT		✓	
Pacose (2020)			✓
UWrMaxSAT		✓	✓

Complete track: Weighted

623 instances

Solver	#Solved	Time (Avg)

Complete track: Weighted

623 instances

Solver	#Solved	Time (Avg)
UWrMaxSAT	455	159.06

Complete track: Weighted

623 instances

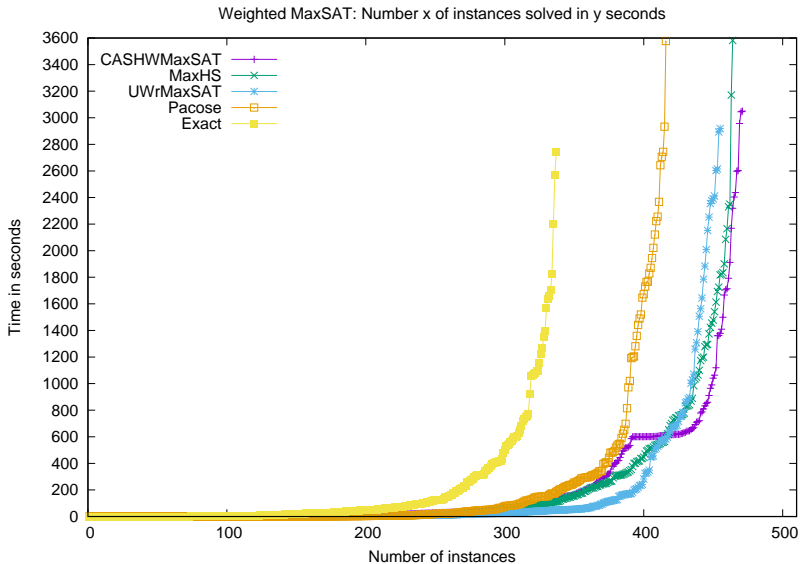
Solver	#Solved	Time (Avg)
MaxHS	464	198.96
UWrMaxSAT	455	159.06

Complete track: Weighted

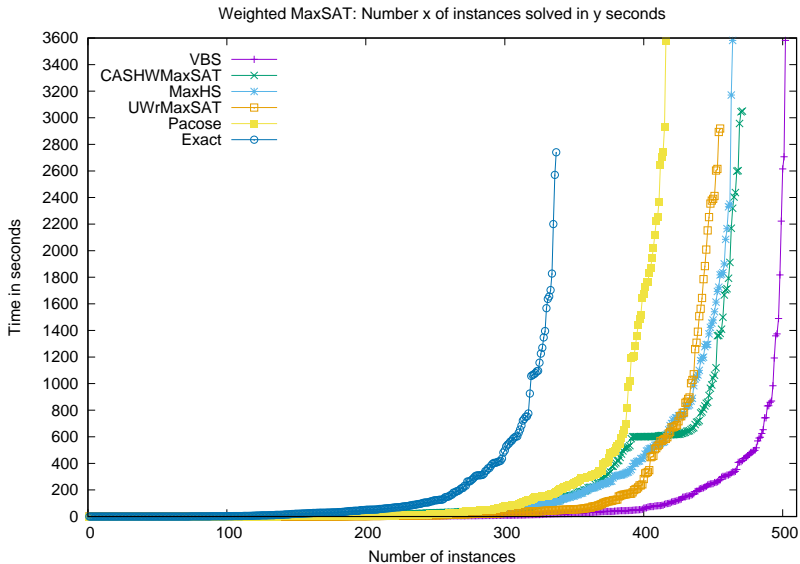
623 instances

Solver	#Solved	Time (Avg)
CASHWMaxSAT	471	217.62
MaxHS	464	198.96
UWrMaxSAT	455	159.06

Complete track: Weighted



Complete track: Weighted



Complete track: Weighted

- ▶ Similar performance with respect to best solvers of recent years:
 - ▶ 2021: CASHWMaxSAT solved 471 instances, 217.62 seconds (average)
 - ▶ 2020: UWrrMaxSAT solved 441 instances, 117.91 seconds (average)
 - ▶ 2018: RC2 solved 428 instances, 203.82 seconds (average)
- ▶ VBS solves 502 instances:
 - ▶ VBS is getting closer to the best solver
 - ▶ Only more 31 instances!
- ▶ Closer look at the VBS:

Pacose	199
UWrMaxSAT	122
MaxHS	115
CASHWMaxSAT	60
Exact	6

Ranking for incomplete tracks

Incomplete ranking:

- ▶ Incomplete score: computed by the sum of the ratios between the best solution found by a given solver and the best known solution:
 - ▶ $\sum_i \frac{(\text{cost of best known solution} + 1)}{(\text{cost of solution for } i \text{ found by solver} + 1)}$
 - ▶ 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 $[0, 1]$

Incomplete Track

Incomplete track: Unweighted

MaxSAT approaches in MSE 2021:

Solver	Stochastic	Unsat-based	Sat-Unsat	Other
Loandra (2020)		✓	✓	✓
StableResolver (2020)	✓			✓
Exact		✓		✓
TT-Open-WBO-Inc	✓		✓	✓
SATLike	✓		✓	

Incomplete track: Unweighted

New and/or improved solvers:

- ▶ **Exact** (new)
 - ▶ Same as complete track
- ▶ **TT-Open-WBO-Inc** by Alexander Nadel.
 - ▶ Uses the SATLike solver as inprocessing.
 - ▶ Includes further modifications to the Polosat algorithm.
 - ▶ More details can be found in the solver description.
- ▶ **SATLike-c** by Zhendong Lei and Shaowei Cai, University of Chinese Academy of Sciences:
 - ▶ Switches between a stochastic algorithm and TT-Open-WBO-Inc.
 - ▶ More details can be found in the solver description.

Incomplete track: Unweighted (60 seconds)

Results ...

Incomplete track: Unweighted (60 seconds)

155 instances

Solver	Score (avg)
Loandra (2020)	0.705
StableResolve (2020)	0.685
Exact	0.422

Incomplete track: Unweighted (60 seconds)

155 instances

Solver	Score (avg)
SATLike-c	0.830
TT-Open-WBO-Inc	0.823
SATLike-ck	0.816
Loandra (2020)	0.705
StableResolve (2020)	0.685
Exact	0.422

- SATLike-c: Hybrid approach between stochastic algorithms and TT-Open-WBO-Inc was the best approach

Incomplete track: Unweighted (300 seconds)

Results ...

Incomplete track: Unweighted (300 seconds)

155 instances

Solver	Score (avg)
Loandra (2020)	0.834
StableResolve (2020)	0.703
Exact	0.470

Incomplete track: Unweighted (300 seconds)

155 instances

Solver	Score (avg)
SATLike-c	0.881
SATLike-ck	0.866
TT-Open-WBO-Inc	0.862
Loandra (2020)	0.834
StableResolve (2020)	0.703
Exact	0.470

- Similar results to 60 seconds

Incomplete track: Weighted

MaxSAT approaches in MSE 2021:

Solver	Stochastic	Unsat-based	Sat-Unsat	Other
Open-WBO-Inc-complete*		✓	✓	✓
Open-WBO-Inc-satlike*	✓	✓		✓
Loandra*		✓	✓	✓
StableResolver*	✓			✓
TT-Open-WBO-Inc	✓		✓	✓
SATLike	✓		✓	
Exact		✓		✓

* denotes solvers from 2020

Incomplete track: Weighted (60 seconds)

Results ...

Incomplete track: Weighted (60 seconds)

151 instances

Solver	Score (avg)
Loandra (2020)	0.728
Open-WBO-Inc-complete (2020)	0.722
Open-WBO-Inc-satlike (2020)	0.684
StableResolve (2020)	0.428
Exact	0.377

Incomplete track: Weighted (60 seconds)

151 instances

Solver	Score (avg)
SATLike-ck	0.793
TT-Open-WBO-Inc	0.785
SATLike-c	0.784
Loandra (2020)	0.728
Open-WBO-Inc-complete (2020)	0.722
Open-WBO-Inc-satlike (2020)	0.684
StableResolve (2020)	0.428
Exact	0.377

- ▶ 3 best solvers use hybrid approaches between stochastic algorithms and SAT-based approaches
- ▶ Memory is becoming an issue for some solvers due to the size of instances (even more this year)

Incomplete track: Weighted (300 seconds)

Results ...

Incomplete track: Weighted (300 seconds)

150 instances

Solver	Score (avg)
SATLike-ck	0.776
Open-WBO-Inc-complete (2020)	0.724
Open-WBO-Inc-satlike (2020)	0.720
StableResolve (2020)	0.583
Exact	0.503

Incomplete track: Weighted (300 seconds)

151 instances

Solver	Score (avg)
Loandra (2020)	0.831
TT-Open-WBO-Inc	0.796
SATLike-c	0.779
SATLike-ck	0.776
Open-WBO-Inc-complete (2020)	0.724
Open-WBO-Inc-satlike (2020)	0.720
StableResolve (2020)	0.583
Exact	0.503

- ▶ Hybrid approaches perform better for larger timeouts
- ▶ The preprocessing employed by Loandra seems to work better with larger timeout.
- ▶ More results available at the MSE website

Webpage

MaxSAT Evaluation 2021 webpage

<https://maxsat-evaluations.github.io/2021/>

- ▶ Tables with average times and number of solved instances
- ▶ Complete ranking tables
- ▶ Cactus plots
- ▶ Detailed results for each instance
- ▶ Description of the solvers
- ▶ Source code of the solvers
- ▶ Description of the benchmarks
- ▶ Benchmarks and log files are available for download

Looking ahead

Format

- ▶ Changes to the format:
 - ▶ No more p-line
 - ▶ No more top
- ▶ Should the format be extended?
 - ▶ Support for floating point weights
 - ▶ Support for negative weights
 - ▶ Support for cardinality constraints
 - ▶ ...

Looking ahead

New tracks

- ▶ Consider other tracks?
 - ▶ Incremental
 - ▶ Enumeration of MCSes
 - ▶ Top-k?
 - ▶ ...

Looking ahead

MaxSAT Lib

<http://www.cs.toronto.edu/maxsat-lib/>

- ▶ Collection of MaxSat instances
- ▶ Make available the best known solution to each instance
- ▶ Community can contribute to update the best known solution
- ▶ Additional resources for MaxSAT research can be made available here

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/>

