

MaxSAT Evaluation 2020

Ruben Martins
CMU

Matti Järvisalo
University Helsinki

Jeremias Berg
University Helsinki

Fahiem Bacchus
University Toronto

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

SAT 2020, July 8, 2020

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 2017-2019:

- ▶ 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

What is new?

- ▶ Compact v-line:
 - ▶ 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
 - ▶ Keep a database of best known solution and update it periodically
- ▶ Top-k track:
 - ▶ Find the best k solutions that satisfy a different set of soft clauses

Evaluation tracks

Evaluation tracks:

- ▶ Unweighted:
 - ▶ No distinction between industrial and crafted benchmarks
- ▶ Weighted:
 - ▶ No distinction between industrial and crafted benchmarks
- ▶ Incomplete:
 - ▶ Two special tracks: unweighted and weighted
- ▶ Top-K (new):
 - ▶ Two special tracks: unweighted and weighted

Execution environment

MSE 2020 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 and Top-K tracks:
 - ▶ 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:
 - ▶ Random selection as in MSE 2019
- ▶ Incomplete track:
 - ▶ Random selection
 - ▶ Hard instances: Only considers instances that cannot be solved optimally in 300 seconds by a collection of complete solvers
- ▶ Top-k track:
 - ▶ Random selection
 - ▶ Easy instances: Only considers instances that can be solved optimally by a collection of complete solvers

New benchmarks

- ▶ Verification of binarized neural networks
- ▶ Maximum probability minimal cut sets
- ▶ Data flow security weaknesses of reconfigurable scan networks
- ▶ Coalition Structure Generation
- ▶ Role-Based Access Control maintenance
- ▶ Synthesis of minimal hardware exploits
- ▶ Most compatible phylogenetic trees over multi-state characters
- ▶ Railway timetabling
- ▶ Single-machine scheduling
- ▶ Analysis of large networks
- ▶ Set covering problems from Italian railways
- ▶ Set covering problems from Steiner Triple Systems
- ▶ Program disambiguation
- ▶ Inference of tumor evolutionary history
- ▶ User authorization query problems

New benchmarks

- ▶ MaxSAT is being used in many applications!
 - ▶ 15 different domains
- ▶ We received a lot of new benchmarks:
 - ▶ **5,091 new benchmarks!**
- ▶ Benchmark size is increasing significantly:
 - ▶ **81 GB** (after gzip compression)
- ▶ MaxSAT solvers can solve very large problems:
 - ▶ More than 20 **million variables** and 130 **million clauses**

MSE 2020 benchmarks

Complete track:

- ▶ Unweighted (576 instances)
- ▶ Weighted (600 instances)

Incomplete track:

- ▶ Unweighted (262 instances)
- ▶ Weighted (253 instances)

Top-K track:

- ▶ Unweighted (73 instances)
- ▶ Weighted (70 instances)

Complete track: Unweighted

MaxSAT approaches in MSE 2020:

Solver	Hitting Set	Unsat-based	Sat-Unsat
EvalMaxSAT		✓	
MaxHS	✓		
Pacose			✓
QMaxSAT			✓
UWrMaxSAT		✓	✓
maxino		✓	
Open-WBO		✓	
smax		✓	
RC2		✓	

- 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
Jeremias Berg, Fahiem Bacchus and Alex Poole: Abstract Cores in Implicit Hitting Set MaxSat Solving. SAT 2020
- ▶ **UWrMaxSAT** by Marek Piotrów, University of Wroclaw
Michal Karpiński and Marek Piotrów: Incremental Encoding of Pseudo-Boolean Goal Functions based on Comparator Networks. SAT 2020
- ▶ **EvalMaxSAT** (new) by Florent Avellaneda, Computer Research Institute of Montreal
Uses similar techniques to RC2.
- ▶ **smax** (new) by Norbert Manthey
Implement a C++ interface as a library that interacts with Open-WBO.

Complete track: Unweighted

576 instances

Solver	#Solved	Time (Avg)

Complete track: Unweighted

576 instances

Solver	#Solved	Time (Avg)
UWrMaxSat	419	163.59

Complete track: Unweighted

576 instances

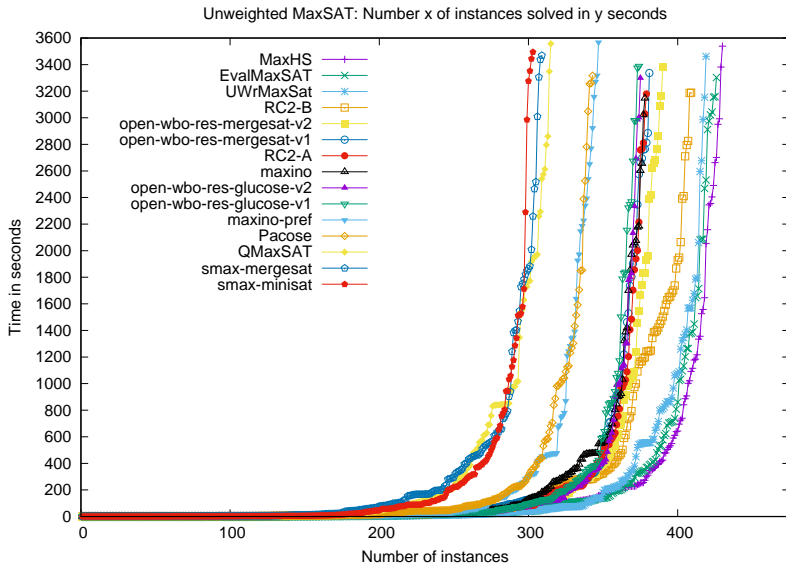
Solver	#Solved	Time (Avg)
EvalMaxSAT	426	182.01
UWrMaxSat	419	163.59

Complete track: Unweighted

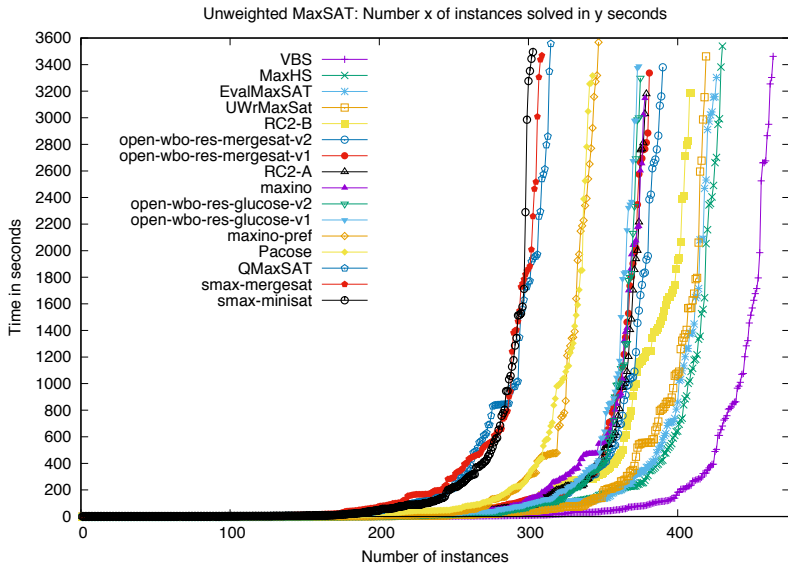
576 instances

Solver	#Solved	Time (Avg)
MaxHS	430	180.16
EvalMaxSAT	426	182.01
UWrMaxSat	419	163.59

Complete track: Unweighted



Complete track: Unweighted



Complete track: Unweighted

- ▶ Improvements with respect to best solver at MSE 2019:
 - ▶ RC2: 409 solved instances, 226.4 seconds (average)
- ▶ VBS solves 464 instances:
 - ▶ Smaller gap than in 2019
- ▶ Large benchmarks can be solved:
 - ▶ ms_200_20_12-2.wcnf:
 - ▶ 1,905,226 variables, 70,138,596 clauses
 - ▶ Solved by Open-WBO-RES-MergeSAT-v2 in 55.13 seconds!
 - ▶ spectre_min.wcnf:
 - ▶ 11,694,886 variables, 41,334,378 clauses
 - ▶ Solved by Pacose in 1514.55 seconds!

Complete track: Weighted

MaxSAT approaches in MSE 2020:

Solver	Hitting Set	Unsat-based	Sat-Unsat
MaxHS	✓		
Pacose			✓
QMaxSAT			✓
UWrMaxSAT		✓	✓
maxino		✓	
smax		✓	
RC2		✓	

Complete track: Weighted

600 instances

Solver	#Solved	Time (Avg)

Complete track: Weighted

600 instances

Solver	#Solved	Time (Avg)
RC2-B	417	205.88

Complete track: Weighted

600 instances

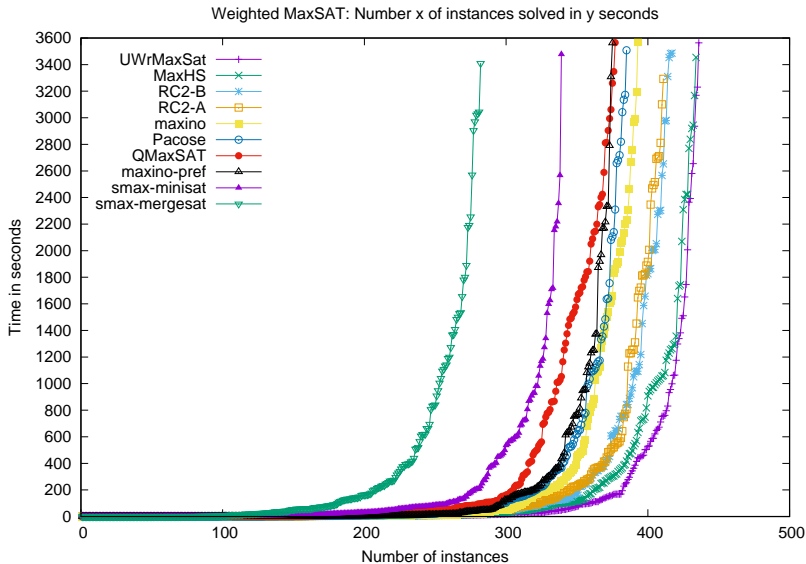
Solver	#Solved	Time (Avg)
MaxHS	434	184.69
RC2-B	417	205.88

Complete track: Weighted

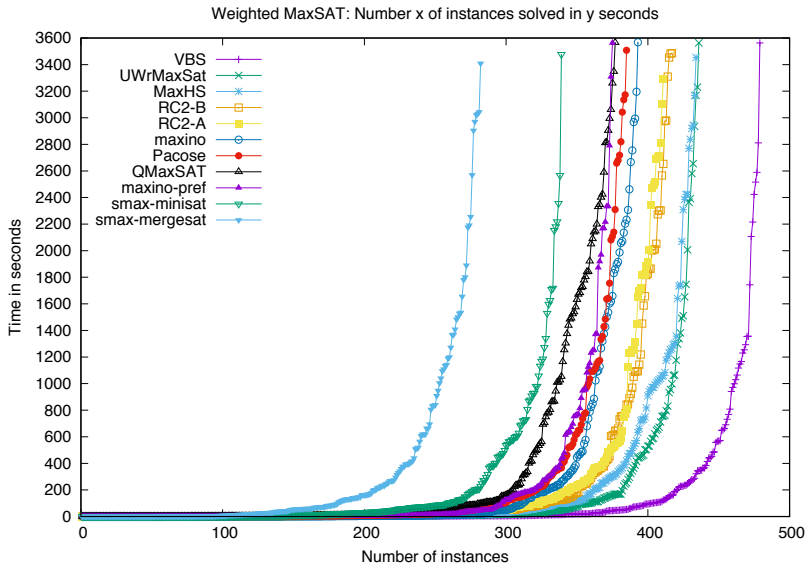
600 instances

Solver	#Solved	Time (Avg)
UWrMaxSat	436	148.97
MaxHS	434	184.69
RC2-B	417	205.88

Complete track: Weighted



Complete track: Weighted



Complete track: Weighted

- ▶ Improvements with respect to best solver at MSE 2019:
 - ▶ RC2: 417 solved instances, 205.88 seconds (average)
- ▶ VBS solves 479 instances:
 - ▶ Smaller gap than in 2019
- ▶ Large benchmarks can be solved:
 - ▶ pa-3.wcnf.gz:
 - ▶ 19,264,629 variables, 22,951,677 clauses
 - ▶ Solved by MaxHS in 731.64 seconds!
 - ▶ bnn_mnist_back_image_73_label5_adversarial_norm_inf_totalizer.wcnf.gz
 - ▶ 1,825,399 variables, 132,682,679 clauses
 - ▶ Solved by Maxino in 2215.51 seconds!

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 solution found by all solvers:
 - ▶ $\sum_i \frac{(\text{cost of best solution for } i \text{ found by any solver} + 1)}{(\text{cost of solution for } i \text{ found by solver} + 1)}$
 - ▶ 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 $[0, 1]$
- ▶ For each instance we consider the best known solution

Incomplete track: Unweighted

MaxSAT approaches in MSE 2020:

Solver	Stochastic	Unsat-based	Sat-Unsat	Other
Loandra		✓	✓	
StableResolver	✓			✓
TT-Open-WBO-Inc			✓	✓
sls-mcs	✓	✓		✓
sls-mcs-lsu	✓	✓		✓
SATLike-c	✓	✓	✓	

Incomplete track: Unweighted

New and/or improved solvers:

- ▶ **StableResolver** (new) by Julian Reisch and Peter Großmann, Synoptics GmbH Dresden, Germany
 - ▶ Gets an initial solution with a SAT solver.
 - ▶ Performs perturbation on the initial solution, improvements and solution checking until the timeout is reached.
 - ▶ More details can be found in the solver description.
- ▶ **TT-Open-WBO-Inc** by Alexander Nadel. On optimizing a generic function in SAT. FMCAD 2020.
- ▶ **SATLike-c** by Zhendong Lei and Shaowei Cai, University of Chinese Academy of Sciences:
 - ▶ Switches between a stochastic algorithm and Loandra.
 - ▶ More details can be found in the solver description.

Incomplete track: Unweighted (60 seconds)

Results ...

Incomplete track: Unweighted (60 seconds)

262 instances

Solver	Score (avg)
sls-mcs-lsu	0.715
sls-mcs	0.701
StableResolver	0.658

Incomplete track: Unweighted (60 seconds)

262 instances

Solver	Score (avg)
SATLike-c	0.798
TT-Open-WBO-Inc	0.79
Loandra	0.738
sls-mcs-lsu	0.715
sls-mcs	0.701
StableResolver	0.658

- ▶ Improvements over last year!
 - ▶ Loandra was the best solver at MSE 2019
- ▶ SATLike-c: Hybrid approach between stochastic algorithms and Loandra was the best approach

Incomplete track: Unweighted (300 seconds)

Results ...

Incomplete track: Unweighted (300 seconds)

262 instances

Solver	Score (avg)
sls-mcs-lsu	0.804
sls-mcs	0.803
StableResolver	0.743

Incomplete track: Unweighted (300 seconds)

262 instances

Solver	Score (avg)
SATLike-c	0.861
TT-Open-WBO-Inc	0.846
Loandra	0.835
sls-mcs-lsu	0.804
sls-mcs	0.803
StableResolver	0.743

- Similar results to 60 seconds

Incomplete track: Weighted

MaxSAT approaches in MSE 2020:

Solver	Stochastic	Unsat-based	Sat-Unsat	Other
Open-WBO-Inc-complete		✓	✓	✓
Open-WBO-Inc-satlike	✓	✓		✓
Open-WBO-Inc-satlike19	✓	✓		✓
Loandra		✓	✓	
StableResolver	✓			✓
TT-Open-WBO-Inc			✓	✓
sls-mcs	✓	✓		✓
sls-mcs2	✓	✓		✓
SATLike-c	✓	✓	✓	
SATLike-cw	✓		✓	✓

Incomplete track: Weighted

New and/or improved solvers:

- ▶ **TT-Open-WBO-Inc** by Alexander Nadel. On optimizing a generic function in SAT. FMCAD 2020.
- ▶ **SATLike-cw** 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: Weighted (60 seconds)

Results ...

Incomplete track: Weighted (60 seconds)

253 instances

Solver	Score (avg)
Loandra	0.741
Open-WBO-Inc-satlike	0.735
SATLike-c	0.735
...	...

Incomplete track: Weighted (60 seconds)

253 instances

Solver	Score (avg)
TT-Open-WBO-Inc	0.808
SATLike-cw	0.802
Open-WBO-Inc-complete	0.769
Loandra	0.741
Open-WBO-Inc-satlike	0.735
SATLike-c	0.735
...	...

- ▶ 3 out of 6 best performing solvers use hybrid approaches between stochastic algorithms and SAT-based approaches
- ▶ Memory is becoming an issue for some instances due to their size

Incomplete track: Weighted (300 seconds)

Results ...

Incomplete track: Weighted (300 seconds)

253 instances

Solver	Score (avg)
Open-WBO-Inc-complete	0.848
Loandra	0.838
Open-WBO-Inc-satlike	0.823
...	...

Incomplete track: Weighted (300 seconds)

253 instances

Solver	Score (avg)
TT-Open-WBO-Inc	0.871
SATLike-cw	0.87
SATLike-c	0.855
Open-WBO-Inc-complete	0.848
Loandra	0.838
Open-WBO-Inc-satlike	0.823
...	...

- ▶ Hybrid approaches perform better for larger timeouts
- ▶ TT-Open-WBO-Inc best performing solver for 60 and 300 seconds
- ▶ More results available at the MSE website

Top-k track

Many applications require enumeration of solutions instead of a single solution!

Find the best k solutions such that:

- ▶ Each solution satisfies the hard clauses
- ▶ Each solution must falsify a different set of soft clauses
- ▶ Each solution must be an optimal solution among the set of solutions falsifying different soft clauses that remain

Score:

- ▶ Ratio between number of models found and k (or existing models if fewer than k)

Setup

Solvers:

- ▶ Open-WBO
- ▶ RC2 (2 versions)
- ▶ maxino (2 versions)
- ▶ MaxHS

Benchmarks:

- ▶ 73 unweighted instances
- ▶ 70 weighted instances

Two settings are evaluated: Top-10 and Top-50

Unweighted

Top-10

Solver	Score
RC2-A	0.9877
RC2-B	0.9781
maxino	0.9740
MaxHS	0.9507
Open-WBO	0.9493
maxino-pref	0.8781

Top-50

Solver	Score
RC2-A	0.9866
maxino	0.9729
RC2-B	0.9655
Open-WBO	0.9460
MaxHS	0.9347
maxino-pref	0.8698

- ▶ RC2 uses maximal satisfiable subset (MSS) enumeration.
- ▶ Most solvers can find all requested solutions for these instances.

Weighted

Top-10

Solver	Score
RC2-B	0.8971
RC2-A	0.8929
MaxHS	0.8914
maxino	0.8900
maxino-pref	0.8114

Top-50

Solver	Score
RC2-B	0.8569
RC2-A	0.8223
MaxHS	0.8220
maxino	0.8177
maxino-pref	0.7537

- ▶ MSS enumeration is currently the best approach
- ▶ Implementing a MaxSAT enumerator was more challenging than expected:
 - ▶ Many optimizations that MaxSAT solvers use are not sound when enumerating solutions

Webpage

MaxSAT Evaluation 2020 webpage

<https://maxsat-evaluations.github.io/2020/>

- ▶ 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 for MSE 2021:
 - ▶ 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

- ▶ Should we continue with the Top-K track?
- ▶ Consider other tracks?
 - ▶ Incremental
 - ▶ Enumeration of MCSes
 - ▶ ...

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/>

