### **MaxSAT Evaluation 2019**

Ruben Martins CMU Matti Järvisalo University Helsinki

Fahiem Bacchus University of Toronto

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

SAT 2019, July 11, 2019

# 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

# **MaxSAT Applications**

- ▶ Many real-world applications can be encoded to MaxSAT:
  - ► Software package upgradeability



► Error localization in C code



Haplotyping with pedigrees



# Outline

► Setup

#### Benchmarks

#### Results

- Complete Tracks
- ► Incomplete Tracks

#### More information

# Setup

Same structure as the one used in MaxSAT Evaluation 2017 and 2018:

- 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

# **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

MSE 2019 did not include a track for random instances!

Contact us if you want to help us revive this track

### **Execution environment**

MSE19 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 track:
  - ► Time limit: 3600 seconds
  - Memory limit: 32 GB
- ► Incomplete track:
  - ► Two time limits: 60 seconds and 300 seconds
  - ► Memory limit: 32 GB

### **Benchmark selection**

This year we implemented a method for selecting the evaluation test set.

#### ► Benchmark pool:

- ▶ Non-random families from all previous evaluations
- 6 new families (2 weighted, 2 unweighted, 2 with both weighted and unweighed instances).
- Filtered out
  - ▶ Very easy instances (solved by 3 different algorithms in 10 sec. or less).
  - ▶ Instances with optimal cost of zero. (These reduce to a SAT problem).
  - For the incomplete tracks, we also tried to remove instances that could be solved exactly in the given time bound.

### **Benchmark selection**

- New procedure. To select N instances for the evaluation suite.
  - ► Select 0.05 × *N* instances from each new family. (20% of the evaluation suite from new families).
  - From the K old families, select  $k_i$  instances from the *i*-th family, where  $k_1, \ldots, k_K$  are random numbers such that  $\sum_i k_i = K$  (multinomial distribution).
  - ► To select *M* instances from a family
    - 1. Measure size of each instance as the sum of the clause sizes (hard and soft).
    - Partition the instances in the family into quintiles based on size (bottom 20% by size to top 20% by size).
    - **3.** Randomly select without replacement  $m_i$  instances from each quintile where  $m_1, ..., m_5$  are 5 random numbers whose sum is M (multinomial distribution).
    - **4.** If family has sub-families use the multinomial distribution to choose how many of the *M* instances are to come from each sub-family then recursively apply the procedure to select from each sub-family.

### New benchmarks

- Minimum Weight Dominating Set Problem (10 benchmarks)
- Identifying Security-Critical Cyber-Physical Components in Weighted AND/OR Graphs (80 benchmarks)
- Consistent Query Answering (19 benchmarks)
- MaxSAT Queries in The Design of Interpretable Rule-based Classifiers (17,135 benchmarks)
- Maximum Common Sub-Graph Extraction (8,544 benchmarks)
- ▶ Parametric RBAC Maintenance via Max-SAT (883 benchmarks)
- Datasets of Networks (8 benchmarks)
- ► 34,626 new benchmarks!

# MSE19 benchmarks

Complete track:

- ► Unweighted (599 benchmarks):
  - ► 48 families of benchmarks
  - ▶ 201 benchmarks were used in MSE18
  - ▶ 66 benchmarks are new
  - ► 332 benchmarks were previously submitted
- ▶ Weighted (586 benchmarks):
  - ► 39 families of benchmarks
  - ▶ 191 benchmarks were used in MSE18
  - ▶ 97 benchmarks are new
  - ▶ 295 benchmarks were previously submitted

# **MSE19** benchmarks

Incomplete track:

- Same selection procedure as the complete track
- ► Unweighted (299 benchmarks):
  - ▶ 112 benchmarks were used in MSE18
  - ▶ 60 benchmarks are new
  - ▶ 127 benchmarks were previously submitted
- ► Weighted (297 benchmarks):
  - ▶ 97 benchmarks were used in MSE18
  - ► 37 benchmarks are new
  - ▶ 163 benchmarks were previously submitted

MaxSAT approaches in MSE19:

Solver	Hitting Set	Unsat-based	Sat-Unsat
maxino		$\checkmark$	
Open-WBO		$\checkmark$	
RC2		$\checkmark$	
UWrMaxSAT		$\checkmark$	
MaxHS	$\checkmark$		
QMaxSAT			$\checkmark$

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

New solvers:

- ► UWrMaxSAT by Marek Piotrów, Institute of Computer Science, University of Wroclaw, Poland.
  - Extends the PB solver kp-minisatp
  - Competitive sorter-based encoding of PB-constraints into SAT. POS 2018.
  - Unsat-based approach (OLL approach)
  - COMiniSatPS as the underlying SAT solver
  - More details in the solver description

Results ...

#### 599 instances

Solver	#Solved	Time (Avg)
maxino2018	399	148.71
Open-WBO-g	399	156.88
Open-WBO-ms-pre	391	151.73
MaxHS	390	182.77
QMaxSAT2018	305	232.95

- ▶ xxx2018 corresponds to the 2018 version of the solver
- ► Open-WBO-g uses **Glucose**
- Open-WBO-ms-pre uses mergesat and the MaxPre preprocessor

#### 599 instances

Solver	#Solved	Time (Avg)
RC2-2018	419	169.44
UWrMaxSAT	414	83.86
Open-WBO-ms	409	174.12
maxino2018	399	148.71
Open-WBO-g	399	156.88
Open-WBO-ms-pre	391	151.73
MaxHS	390	182.77
QMaxSAT2018	305	232.95

- Open-WBO-ms uses mergesat
- ► UWrMaxSAT is faster than other solvers
- ▶ Not many improvements with respect to the best solvers from 2018

RC2-2018 (best solver) solves 419 benchmarks VBS solves 467 benchmarks!

Solver	#Solved in VBS
UWrMaxSAT	108
maxino2018	89
Open-WBO-g	83
QMaxSAT2018	65
MaxHS	55
RC2-2018	35
Open-WBO-ms-pre	19
Open-WBO-ms	13

Unweighted MaxSAT: Number x of instances solved in y seconds



Unweighted MaxSAT: Number x of instances solved in y seconds



MaxSAT approaches in MSE18:

Solver	Hitting Set	Unsat-based	Sat-Unsat
maxino		$\checkmark$	
Open-WBO		$\checkmark$	
RC2		$\checkmark$	
UWrMaxSAT		$\checkmark$	
MaxHS	$\checkmark$		
QMaxSAT			$\checkmark$
Pacose			$\checkmark$

Same solvers as in the unweighted track plus Pacose

Results . . .

#### 586 instances

Solver	#Solved	Time (Avg)
QMaxSAT2018	327	321.78
maxino2018	325	221.59
Pacose	321	309.23
Open-WBO-g	317	277.12
Open-WBO-ms-pre	311	191.96
Open-WBO-ms	306	264.3

Preprocessing can help:

▶ Integration between solver and preprocessor can be improved

Underlying SAT solver can have an impact

#### 586 instances

Solver	#Solved	Time (Avg)
RC2-2018	380	269.58
UWrMaxSAT	371	186.33
MaxHS	357	259.52
QMaxSAT2018	327	321.78
maxino2018	325	221.59
Pacose	321	309.23
Open-WBO-g	317	277.12
Open-WBO-ms-pre	311	191.96
Open-WBO-ms	306	264.3

▶ RC2 wins both unweighted and weighted tracks (like in 2018)!

RC2-2018 (best solver) solves 380 benchmarks VBS solves 459 benchmarks!

Solver	#Solved in VBS
MaxHS	98
UWrMaxSAT	78
maxino2018	67
Open-WBO-g	83
QMaxSAT2018	38
Pacose	51
RC2-2018	20
Open-WBO-ms-pre	12
Open-WBO-ms	12

Unweighted MaxSAT: Number x of instances solved in y seconds



Unweighted MaxSAT: Number x of instances solved in y 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:

 $\blacktriangleright \sum_{i} \frac{(\text{cost of best solution for i found by any solver} + 1)}{(\text{cost of solution for i found by solver} + 1)}$ 

- ▶ 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]
- ► For each instance we consider the best solution found by all incomplete solvers within 300 seconds

MaxSAT approaches in MSE19:

Solver	Stochastic	Unsat-based	Sat-Unsat	Other
LinSBPS2018			$\checkmark$	
Open-WBO-g		$\checkmark$		
Open-WBO-ms		$\checkmark$		
SATLike-c	$\checkmark$		$\checkmark$	
Loandra		$\checkmark$	$\checkmark$	
sls-mcs	$\checkmark$	$\checkmark$		$\checkmark$
sls-mcs-lsu	$\checkmark$	$\checkmark$		$\checkmark$

► New approaches for incomplete MaxSAT!

New solvers:

- Loandra by Jeremias Berg (University of Helsinki, Finland), Emir Demirović and Peter J. Stuckey (University of Melbourne, Australia).
  - Switches between core-guided and model improving algorithms.
  - ► Core-boosted linear search for incomplete MaxSAT. CPAIOR 2019
  - More details in the paper and solver description
- sls by Andreia P. Guerreiro, Miguel Terra-Neves, Ines Lynce, Jose Rui Figueira, and Vasco Manquinho (Instituto Superior Tecnico, Universidade de Lisboa, Portugal).
  - Integrates SAT-based techniques in a Stochastic Local Search solver for MaxSAT.
  - ▶ More details can be found in the CP 2019 paper and solver description

Results ...

299 instances

Solver	Score (avg)
sls-mcs-lsu	0.683
sls-mcs	0.683
Open-WBO-ms	0.606

#### 299 instances

Solver	Score (avg)
Loandra	0.809
LinSBPS2018	0.779
Open-WBO-g	0.689
sls-mcs-lsu	0.683
sls-mcs	0.683
Open-WBO-ms	0.606

- Improvements over last year!
- Some solvers cannot find any solutions on many benchmarks:
  - ► Loandra cannot find at least one solution on 16 benchmarks
  - ▶ Open-WBO-ms cannot find at least one solution on **43 benchmarks**
  - sls-mcs cannot find at least one solution on 70 benchmarks!

299 instances

Solver	Score (avg)
Loandra	0.809
LinSBPS2018	0.779
SATLike*	0.771
Open-WBO-g	0.689
sls-mcs-lsu	0.683
sls-mcs	0.683
Open-WBO-ms	0.606

\*SATLike reported assignments that do not satisfy hard clauses:

- ▶ It could be a printing error and the 'o' lines still be correct.
- Score uses the VBS without SATLike (i.e., some instances SATLike will have score > 1.0).
- Scoring metric is dependent on the solvers used for the best solution. Suggestions for other metrics?

Results ...

299 instances

Solver	Score (avg)
sls-mcs-lsu	0.782
Open-WBO-g	0.749
Open-WBO-ms	0.730

299 instances

Solver	Score (avg)
Loandra	0.884
LinSBPS2018	0.847
sls-mcs	0.797
sls-mcs-lsu	0.782
Open-WBO-g	0.749
Open-WBO-ms	0.730

- Loandra is the winner for both time limits!
- ▶ Improvements over last year on both 60 and 300 seconds!

#### 299 instances

Solver	Score (avg)
Loandra	0.884
SATLike*	0.860
LinSBPS2018	0.847
sls-mcs	0.797
sls-mcs-lsu	0.782
Open-WBO-g	0.749
Open-WBO-ms	0.730

\*SATLike reported assignments that do not satisfy hard clauses:

- ▶ It could be a printing error and the 'o' lines still be correct.
- Score uses the VBS without SATLike (i.e., some instances SATLike will have score > 1.0).

MaxSAT approaches in MSE19:

Solver	Stochastic	Unsat-based	Sat-Unsat	Other
LinSBPS2018			$\checkmark$	
Open-WBO-g		$\checkmark$		
Open-WBO-ms		$\checkmark$		
SATLike-c	$\checkmark$		$\checkmark$	
Loandra		$\checkmark$	$\checkmark$	
sls-mcs	$\checkmark$	$\checkmark$		$\checkmark$
sls-mcs2	$\checkmark$	$\checkmark$		$\checkmark$
TT-Open-WBO-Inc		$\checkmark$		
Open-WBO-Inc-bc		$\checkmark$	$\checkmark$	
Open-WBO-Inc-bs	$\checkmark$	$\checkmark$		
uwrmaxsat-inc		$\checkmark$		

► New approaches for incomplete MaxSAT!

New solvers:

- TT-Open-WBO-Inc by Alexander Nadel (Intel, Israel).
  - Polarity selection heuristic (TORC)
  - Enhancement to the variable selection strategy (TSB)
  - Paper under submission
  - More details in the solver description

Results ...

297 instances

Solver	Score (avg)
Open-WBO-Inc-bc	0.738
LinSBPS2018	0.726
Open-WBO-g	0.715
sls-mcs2	0.685
Open-WBO-ms	0.656
sls-mcs	0.646
uwrmaxsat-inc	0.643

297 instances

Solver	Score (avg)
tt-open-wbo-inc	0.860
Loandra	0.843
Open-WBO-Inc-bs	0.827
Open-WBO-Inc-bc	0.738
LinSBPS2018	0.726
Open-WBO-g	0.715
sls-mcs2	0.685
Open-WBO-ms	0.656
sls-mcs	0.646
uwrmaxsat-inc	0.643

- Improvements over last year!
- Some solvers cannot find any solutions on many benchmarks:
  - tt-open-wbo-inc: 22 benchmarks
  - sls-mcs: 56 benchmarks
  - uwrmaxsat-inc: 59 benchmarks

#### 297 instances

Solver	Score (avg)
tt-open-wbo-inc	0.860
Loandra	0.843
Open-WBO-Inc-bs	0.827
Open-WBO-Inc-bc	0.738
LinSBPS2018	0.726
Open-WBO-g	0.715
SATLike*	0.708
sls-mcs2	0.685
Open-WBO-ms	0.656
sls-mcs	0.646
uwrmaxsat-inc	0.643

\*SATLike reported assignments that do not satisfy hard clauses

Results ...

297 instances

Solver	Score (avg)
LinSBPS2018	0.823
Open-WBO-Inc-bc	0.815
Open-WBO-g	0.788
sls-mcs2	0.746
uwrmaxsat-inc	0.741
Open-WBO-ms	0.736
sls-mcs	0.698

#### 297 instances

Solver	Score (avg)
tt-open-wbo-inc	0.860
Loandra	0.843
Open-WBO-Inc-bs	0.827
LinSBPS2018	0.823
Open-WBO-Inc-bc	0.815
Open-WBO-g	0.788
sls-mcs2	0.746
uwrmaxsat-inc	0.741
Open-WBO-ms	0.736
sls-mcs	0.698

**tt-open-wbo-inc** is the best solver for both time limits!

#### 297 instances

Solver	Score (avg)
tt-open-wbo-inc	0.860
Loandra	0.843
Open-WBO-Inc-bs	0.827
LinSBPS2018	0.823
SATLike*	0.772
Open-WBO-Inc-bc	0.815
Open-WBO-g	0.788
sls-mcs2	0.746
uwrmaxsat-inc	0.741
Open-WBO-ms	0.736
sls-mcs	0.698

\*SATLike reported assignments that do not satisfy hard clauses

# Webpage

#### MaxSAT Evaluation 2019 webpage

https://maxsat-evaluations.github.io/2019/

- 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

#### Incomplete track

- Before MaxSAT Evaluation 2017, the organizers were using the number of times a solver found the best solution as the ranking metric
- In the last 2 years, we used the score as a ranking metric. This gives a ratio of how far on average each solver is from the best solution
- Should we use other metrics that are not dependent on the solvers being tested?
  - Send your suggestions to the organizers!

# Looking ahead

#### Incremental MaxSAT solving

- ► A lot of people are starting to ask if current MaxSAT solvers support incremental changes after an optimum solution has been found!
- ▶ Should we create a track for incremental MaxSAT solving?
- Solvers need to be able to simulate:
  - Addition/deletion of hard clauses
  - Addition/deletion of soft clauses
- We need to discuss on a common interface that all solvers will need to support. Suggestions?

# Looking ahead

#### Challenge problems

- Earlier in this conference, we discussed challenge problems for SAT!
- Should we submit MaxSAT challenge problems?
  - Easier to keep track of progress (improvements on lower and upper bounds).
- What kind of problems would be interesting?

#### Single domain problems

Should we have a track in the MSE20 for single domains? MaxSAT Queries in The Design of Interpretable Rule-based Classifiers, Maximum Common Sub-Graph Extraction, other?

#### 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/

