#### MaxSAT Evaluation 2018

Ruben Martins CMU Matti Järvisalo University Helsinki

Fahiem Bacchus University of Toronto

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

SAT 2018, July 12, 2018

# MaxSAT (R)evolution - Unweighted

MSE17 benchmarks:



# MaxSAT (R)evolution - Weighted

MSE17 benchmarks:



Evolution of Weighted MaxSAT Solvers

#### Outline

- ► Setup
- Benchmarks
- Results
  - ► Complete Tracks
  - ► Incomplete Tracks
- ► More information

# Setup

A lot has changed in the MaxSAT Evaluation 2017 (MSE17). This year we used the same structure as the one used in the MSE17:

- ► 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 2018 did not include a track for random instances!

#### **Execution environment**

MSE18 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**

Complete benchmarks:

- ► Benchmark pool:
  - All MSE17 benchmarks
  - ► All new benchmarks submitted to MSE17
  - ► All new benchmarks submitted to MSE18
- Random selection:
  - ▶ Maximum 25 instances for older benchmark sets (MSE17)
  - ▶ Maximum 40 instances for new benchmark sets (MSE18)
  - ► Instances selected randomly from the pool of benchmarks

Incomplete benchmarks:

- ► Hard benchmarks:
  - Only consider the subset of benchmarks that are not solved optimally under 300 seconds

### New benchmarks

Unweighted (351 new benchmarks):

- ► drmx-atmostk (36)
- ► drmx-cryptogen (40)
- ▶ optic (65)
- ▶ uaq (97)
- ▶ vpa (67)
- ▶ xai-mindset (46)

Weighted (244 new benchmarks):

- drmx-atmostk (36)
- drmx-cryptogen (40)
- ▶ tcp (60)
- ▶ cluster-expansion (21)
- ▶ power-distribution-full (28)
- ▶ power-distribution-sparse (56)
- ▶ robot-navigation (3)

# MSE18 benchmarks

Complete track:

- Unweighted (600 benchmarks):
  - $\blacktriangleright$  66% of the benchmarks were used in MSE17
  - ▶ 34% of the benchmarks are new
- Weighted (600 benchmarks):
  - $\blacktriangleright$  65% of the benchmarks were used in MSE17
  - $\blacktriangleright~35\%$  of the benchmarks are new

Incomplete track:

(selection of benchmarks that complete solvers take more than 300 seconds to find the optimal solution or that no optimal solution is found)

- ▶ Unweighted (153 benchmarks)
- ▶ Weighted (172 benchmarks)

MaxSAT approaches in MSE18:

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

- ▶ Branch & Bound is no longer being used by any solver!
- ► Diverse approaches in MaxSAT!
- ► Each approach is important and can solve different applications!

New solvers:

- ► RC2 by Alexey Ignatiev, Antonio Morgado, and Joao Marques-Silva, Faculty of Sciences, University of Lisbon, Portugal.
  - ▶ PySAT: A Python Toolkit for Prototyping with SAT Oracles. SAT 2018
  - Unsat-based approach
  - ► RC2-A: no core minimization
  - ► RC2-B: limited core minimization
  - More details in the solver description

Results . . .

#### 600 instances

Solver	#Solved	Time (Avg)
RC2-B	421	126.32
RC2-A	416	138.98
maxino	405	137.5
MaxHS	386	178.06
Open-WBO-Gluc	382	171.54

- ▶ Similar version to Open-WBO-Gluc was the best solver in MSE17
- Comparison with MSE17:
  - ▶ 39 more benchmarks solved!
  - Note: Different approaches solve different problems and this may change from year to year. Example: Last year maxino solved 13 less benchmarks than Open-WBO and this year it solves more 23!

RC2-B (best solver) solves 421 benchmarks VBS solves 472 benchmarks!

► Contribution to the VBS:



Note: there are more solvers using an Unsat-based approach

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		$\checkmark$	
LMHS	$\checkmark$		
MaxHS	$\checkmark$		
QMaxSAT			$\checkmark$
Pacose			$\checkmark$

► Same solvers as in the unweighted track plus Pacose

New solvers:

- RC2 by Alexey Ignatiev, Antonio Morgado, and Joao Marques-Silva, Faculty of Sciences, University of Lisbon, Portugal.
- ► **Pacose** by Tobias Paxian, Sven Reimer, and Bernd Becker, Albert-Ludwigs-Universität Freiburg, Germany.
  - ► Linear search Sat-Unsat with a new PB encoding
  - Built on top of QMaxSAT
  - Dynamic polynomial watchdog encoding for solving weighted MaxSAT. SAT 2018
  - ▶ More details in the SAT paper and solver description

Results . . .

#### 600 instances

Solver	#Solved	Time (Avg)
RC2-B	421	256.02
RC2-A	416	267.55
MaxHS	390	274.87
Pacose	390	348.98
QMaxSAT	381	320.78

- ► MaxHS was the best solver in MSE17
- Comparison with MSE17:
  - ▶ 31 more benchmarks solved!

RC2-B (best solver) solves 421 benchmarks VBS solves 499 benchmarks!

► Contribution to the VBS:



Note: there are more solvers using an Unsat-based approach

Weighted MaxSAT: Number x of instances solved in y seconds



Weighted 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 MSE18:

Solver	Stochastic	Unsat-based	Sat-Unsat	Other
LinSBPS			$\checkmark$	
maxroster	$\checkmark$	$\checkmark$	$\checkmark$	
Open-WBO			$\checkmark$	
Open-WBO-Inc-OBV			$\checkmark$	$\checkmark$
Open-WBO-Inc-MCS			$\checkmark$	$\checkmark$
SATLike	$\checkmark$			
SATLike-c	$\checkmark$		$\checkmark$	

- ▶ No Hitting Set approaches submitted to the incomplete track
- ► New approaches for incomplete MaxSAT!

New solvers:

- LinSBPS by Emir Demirović and Peter J. Stuckey, University of Melbourne, Australia.
  - Local-Style Search in the Linear MaxSAT Algorithm: A Computational Study of Solution-Based Phase Saving. POS 2018
  - Solution-based phase saving
  - More details in the paper and solver description
- ► Open-WBO-Inc-OBV by Ruben Martins (CMU, USA), Saurabh Joshi, Prateek Kumar, Sukrut Rao (IIT-Hyderabad, India), Vasco Manquinho (INESC-ID, Portugal), Alexander Nadel (Intel, Israel).
  - ► Solving MaxSAT with Bit-Vector Optimization. SAT 2018
  - ► Uses an incomplete Bit-Vector Optimization approach
  - More details in the solver description

New solvers:

- Open-WBO-Inc-MCS by Ruben Martins (CMU, USA), Saurabh Joshi, Prateek Kumar, Sukrut Rao (IIT-Hyderabad, India), Vasco Manquinho (INESC-ID, Portugal), Alexander Nadel (Intel, Israel).
  - ► Finds solution by MCS enumeration
  - More details in the solver description
- ► **SATLike** by Zhendong Lei and Shaowei Cai, Institute of Software Chinese Academy of Sciences, Beijing, China.
  - From Decimation to Local Search and Back: A New Approach to MaxSAT. IJCAI 2017
  - Stochastic search for MaxSAT
  - More details in the solver description
- SATLike-c by Zhendong Lei and Shaowei Cai, Institute of Software Chinese Academy of Sciences, Beijing, China.
  - Initial stochastic search for MaxSAT
  - ▶ Switches to Sat-Unsat algorithm if no solution is found in 50 seconds
  - More details in the solver description

Results ...

153 instances

Solver	Score (avg)
SATLike-c	0.735
LinSBPS	0.705
SATLike	0.675
Open-WBO-Inc-OBV	0.654
Open-WBO-Inc-MCS	0.631

- Stochastic search performing well for incomplete MaxSAT!
- Combining stochastic with Sat-Unsat leads to the best result!
- ► Other approaches based on OBV and MCSes can also be useful
- Comparison with MSE17:
  - ▶ Open-WBO-Gluc (similar to Open-WBO-LSU in MSE17): 0.612
  - maxroster: 0.541
  - ► All new approaches improve previous approaches on 60 seconds!

153 instances

Solver	Score (avg)
SATLike-c	0.854
maxroster	0.829
LinSBPS	0.782
SATLike	0.718
Open-WBO-Inc-OBV	0.713

- ► SATLike-c is the best overall approach for unweighted incomplete
- maxroster is much better for 300 seconds

Score

Incomplete Unweighted MaxSAT (300s): distribution of scores per instances



MaxSAT approaches in MSE18:

Solver	Stochastic	Unsat	Sat-Unsat	Weight-Relax
LinSBPS			$\checkmark$	$\checkmark$
maxroster	$\checkmark$	$\checkmark$	$\checkmark$	
Open-WBO			$\checkmark$	
Open-WBO-Inc-BMO			$\checkmark$	$\checkmark$
Open-WBO-Inc-Cluster			$\checkmark$	$\checkmark$
SATLike	$\checkmark$			
SATLike-c	$\checkmark$		$\checkmark$	

▶ New approaches for incomplete weighted MaxSAT!

New solvers:

- LinSBPS by Emir Demirović and Peter J. Stuckey, University of Melbourne, Australia.
  - Local-Style Search in the Linear MaxSAT Algorithm: A Computational Study of Solution-Based Phase Saving. POS 2018
  - Solution-based phase saving
  - ► Varying resolution approach (more details in the solver description)
- Open-WBO-Inc-BMO by Ruben Martins (CMU, USA), Saurabh Joshi, Prateek Kumar, Sukrut Rao (IIT-Hyderabad, India), Vasco Manquinho (INESC-ID, Portugal), Alexander Nadel (Intel, Israel).
  - ► Approximation Strategies for Incomplete MaxSAT. CP 2018
  - ► Considers each weight as lexicographical optimization function
  - More details in the paper and solver description

New solvers:

- ▶ Open-WBO-Inc-Cluster by Ruben Martins (CMU, USA), Saurabh Joshi, Prateek Kumar, Sukrut Rao (IIT-Hyderabad, India), Vasco Manquinho (INESC-ID, Portugal), Alexander Nadel (Intel, Israel).
  - ► Approximation Strategies for Incomplete MaxSAT. CP 2018
  - Performs weight relaxation
  - More details in the paper and solver description
- ► **SATLike** by Zhendong Lei and Shaowei Cai, Institute of Software Chinese Academy of Sciences, Beijing, China.
- SATLike-c by Zhendong Lei and Shaowei Cai, Institute of Software Chinese Academy of Sciences, Beijing, China.

Results ...

172 instances

Solver	Score (avg)
Open-WBO-Inc-BMO	0.810
LinSBPS	0.799
maxroster	0.773
Open-WBO-Inc-Cluster	0.743
SATLike-c	0.696

- ► New weight-relaxation approaches outperform previous approaches
- ► Stochastic search not as efficient as in unweighted

172 instances

Solver	Score (avg)
LinSBPS	0.900
Open-WBO-Inc-BMO	0.842
maxroster	0.804
Open-WBO-Inc-Cluster	0.762
SATLike-c	0.747

- ► New weight-relaxation approaches outperform previous approaches
- LinSBPS outperforms Open-WBO-Inc-BMO on 300 seconds by doing a gradual weight-relaxation approach

Score

Incomplete Weighted MaxSAT (300s): distribution of scores per instances



# Webpage

#### MaxSAT Evaluation 2018 webpage

https://maxsat-evaluations.github.io/2018/

- ► 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
- ► SQLite database with all results

## Looking ahead

#### Benchmarks

- Create a large library with all the available benchmarks
- ► Random selection on all benchmarks:
  - Reduce possible biases
  - ▶ Reduce the number of benchmarks that intersect with previous year
- This year we decreased the benchmark set to 600 instances. Should we increase it for next year?
- Benchmarks are always welcome! If you work on MaxSAT, do not forget to submit your benchmarks next year!

# 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?
  - ► Primal integral?
  - Consider all intermediate solutions found by incomplete solvers and compute the underlying area limited by these solutions
  - ► Gives higher score to solvers that find better solutions quickly
  - 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?
- If you have benchmarks for incremental MaxSAT, please send us an email to see if there is enough traction to start this track!

#### Thanks

Thanks to the people that contributed solvers and benchmarks:

Mario Alviano, Fahiem Bacchus, Wenxuan Huang, Matti Järvisalo, Ruben Martins, Paul Saikko, Takayuki Sugawara, Alexey Ignatiev, Oleg Zaikin, Rayna Dimitrova, Rüdiger Ehlers, Matthias Heizmann, Christian Schilling, Alessandro Armando, Giorgia Gazzarata, Joao Marques-Silva, Emir Demirović, Peter J. Stuckey, Saurabh Joshi, Norbert Manthey, Tobias Paxian, Aolong Zha, Zhendong Lei, Shaowei Cai

Thanks to StarExec for allowing us to use their cluster:

https://www.starexec.org/

