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:

Evolution of Unweighted MaxSAT Solvers

- Open-WBO (2015)
- Open-WBO (2017)
- MaxHS (2017)
- MaxHS (2016)
- maxino (2015)
- eva (2014)
- Open-WBO (2014)
- QMaxSAT (2013)
- PM2 (2010)
- MSUnCore (2013)
- PWBO (2012)
- QMaxSAT (2011)
- QMaxSAT (2010)
- SAT4J (2009)
- IncWMaxSatz (2008)

CPU Time (in seconds)

Number of problems solved

Years:
- 2008
- 2009
- 2010
- 2011-2012
- 2013
- 2014-2017
MaxSAT (R)evolution - Weighted
MSE17 benchmarks:

Evolution of Weighted MaxSAT Solvers

- **MaxHS (2017)**
- **MaxHS (2016)**
- **maxino (2017)**
- **QMaxSAT (2017)**
- **MSCG (2015)**
- **maxino (2015)**
- **QMaxSAT (2014)**
- **eva (2014)**
- **MaxHS (2013)**
- **WPM2 (2013)**
- **WPM1 (2012)**
- **WBO (2010)**
- **WPM1 (2011)**
- **IncWMaxSatz (2008)**
- **SAT4J (2009)**

CPU Time (in seconds)
Number of problems solved

- 2009-2011
- 2012
- 2013-2014
- 2015-2017
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/](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):
  - 66% of the benchmarks were used in MSE17
  - 34% of the benchmarks are new
- Weighted (600 benchmarks):
  - 65% of the benchmarks were used in MSE17
  - 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)
Complete track: Unweighted

MaxSAT approaches in MSE18:

<table>
<thead>
<tr>
<th>Solver</th>
<th>Hitting Set</th>
<th>Unsat-based</th>
<th>Sat-Unsat</th>
</tr>
</thead>
<tbody>
<tr>
<td>maxino</td>
<td></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Open-WBO</td>
<td></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>RC2</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>LMHS</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>MaxHS</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QMaxSAT</td>
<td></td>
<td></td>
<td>✔</td>
</tr>
</tbody>
</table>

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

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
Complete track: Unweighted

Results ...
### Complete track: Unweighted

600 instances

<table>
<thead>
<tr>
<th>Solver</th>
<th>#Solved</th>
<th>Time (Avg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC2-B</td>
<td>421</td>
<td>126.32</td>
</tr>
<tr>
<td>RC2-A</td>
<td>416</td>
<td>138.98</td>
</tr>
<tr>
<td>maxino</td>
<td>405</td>
<td>137.5</td>
</tr>
<tr>
<td>MaxHS</td>
<td>386</td>
<td>178.06</td>
</tr>
<tr>
<td>Open-WBO-Gluc</td>
<td>382</td>
<td>171.54</td>
</tr>
</tbody>
</table>

- 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!
Complete track: Unweighted

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
Complete track: Unweighted

Unweighted MaxSAT: Number x of instances solved in y seconds

- RC2-B
- RC2-A
- maxino
- MaxHS
- Open-WBO-Gluc
- Open-WBO-Riss
- LMHS
- QMaxSAT

Time in seconds vs. Number of instances solved.
Complete track: Unweighted

Unweighted MaxSAT: Number $x$ of instances solved in $y$ seconds

- VBS
- RC2-B
- RC2-A
- maxino
- MaxHS
- Open-WBO-Gluc
- Open-WBO-Riss
- LMHS
- QMaxSAT

![Graph showing the number of instances solved in seconds for different solvers](image-url)
Complete track: Weighted

MaxSAT approaches in MSE18:

<table>
<thead>
<tr>
<th>Solver</th>
<th>Hitting Set</th>
<th>Unsat-based</th>
<th>Sat-Unsat</th>
</tr>
</thead>
<tbody>
<tr>
<td>maxino</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Open-WBO</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC2</td>
<td></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>LMHS</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MaxHS</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QMaxSAT</td>
<td></td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>Pacose</td>
<td></td>
<td></td>
<td>✔</td>
</tr>
</tbody>
</table>

- Same solvers as in the unweighted track plus Pacose
Complete track: Weighted

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
Complete track: Weighted

Results . . .
Complete track: Weighted

600 instances

<table>
<thead>
<tr>
<th>Solver</th>
<th>#Solved</th>
<th>Time (Avg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC2-B</td>
<td>421</td>
<td>256.02</td>
</tr>
<tr>
<td>RC2-A</td>
<td>416</td>
<td>267.55</td>
</tr>
<tr>
<td>MaxHS</td>
<td>390</td>
<td>274.87</td>
</tr>
<tr>
<td>Pacose</td>
<td>390</td>
<td>348.98</td>
</tr>
<tr>
<td>QMaxSAT</td>
<td>381</td>
<td>320.78</td>
</tr>
</tbody>
</table>

▶ MaxHS was the best solver in MSE17
▶ Comparison with MSE17:
  ▶ 31 more benchmarks solved!
Complete track: Weighted

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
Complete track: Weighted

Weighted MaxSAT: Number $x$ of instances solved in $y$ seconds

RC2-B
RC2-A
MaxHS
Pacose
QMaxSAT
maxino
Open-WBO-Gluc
Open-WBO-Riss
LMHS
Complete track: Weighted

Weighted MaxSAT: Number x of instances solved in y seconds

- VBS
- RC2-B
- RC2-A
- MaxHS
- Pacose
- QMaxSAT
- maxino
- Open-WBO-Gluc
- Open-WBO-Riss
- LMHS

Time in seconds vs. Number of instances
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 solution found by all incomplete solvers within 300 seconds
## Incomplete track: Unweighted (60 seconds)

MaxSAT approaches in MSE18:

<table>
<thead>
<tr>
<th>Solver</th>
<th>Stochastic</th>
<th>Unsat-based</th>
<th>Sat-Unsat</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>LinSBPS</td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>maxroster</td>
<td>✔</td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Open-WBO</td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Open-WBO-Inc-OBV</td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Open-WBO-Inc-MCS</td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>SATLike</td>
<td>✔</td>
<td></td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>SATLike-c</td>
<td>✔</td>
<td></td>
<td></td>
<td>✔</td>
</tr>
</tbody>
</table>

- No Hitting Set approaches submitted to the incomplete track
- New approaches for incomplete MaxSAT!
Incomplete track: Unweighted (60 seconds)

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
Incomplete track: Unweighted (60 seconds)

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
Incomplete track: Unweighted (60 seconds)

Results . . .
Incomplete track: Unweighted (60 seconds)

153 instances

<table>
<thead>
<tr>
<th>Solver</th>
<th>Score (avg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SATLike-c</td>
<td>0.735</td>
</tr>
<tr>
<td>LinSBPS</td>
<td>0.705</td>
</tr>
<tr>
<td>SATLike</td>
<td>0.675</td>
</tr>
<tr>
<td>Open-WBO-Inc-OBV</td>
<td>0.654</td>
</tr>
<tr>
<td>Open-WBO-Inc-MCS</td>
<td>0.631</td>
</tr>
</tbody>
</table>

▶ 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!
Incomplete track: Unweighted (300 seconds)

153 instances

<table>
<thead>
<tr>
<th>Solver</th>
<th>Score (avg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SATLike-c</td>
<td>0.854</td>
</tr>
<tr>
<td>maxroster</td>
<td>0.829</td>
</tr>
<tr>
<td>LinSBPS</td>
<td>0.782</td>
</tr>
<tr>
<td>SATLike</td>
<td>0.718</td>
</tr>
<tr>
<td>Open-WBO-Inc-OBV</td>
<td>0.713</td>
</tr>
</tbody>
</table>

- SATLike-c is the best overall approach for unweighted incomplete
- maxroster is much better for 300 seconds
Incomplete track: Unweighted (300 seconds)

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

Score

Instances

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170

SATLike-c
maxroster
LinSBPS
SATLike
Open-WBO-Inc-OBV
Open-WBO-Inc-MCS
Open-WBO-Gluc
Open-WBO-Riss

15 / 20
### Incomplete track: Weighted (60 seconds)

MaxSAT approaches in MSE18:

<table>
<thead>
<tr>
<th>Solver</th>
<th>Stochastic</th>
<th>Unsat</th>
<th>Sat-Unsat</th>
<th>Weight-Relax</th>
</tr>
</thead>
<tbody>
<tr>
<td>LinSBPS</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>maxroster</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Open-WBO</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Open-WBO-Inc-BMO</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Open-WBO-Inc-Cluster</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>SATLike</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SATLike-c</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- New approaches for incomplete weighted MaxSAT!
Incomplete track: Weighted (60 seconds)

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
Incomplete track: Weighted (60 seconds)

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.
Incomplete track: Weighted (60 seconds)

Results ...
Incomplete track: Weighted (60 seconds)

172 instances

<table>
<thead>
<tr>
<th>Solver</th>
<th>Score (avg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open-WBO-Inc-BMO</td>
<td>0.810</td>
</tr>
<tr>
<td>LinSBPS</td>
<td>0.799</td>
</tr>
<tr>
<td>maxroster</td>
<td>0.773</td>
</tr>
<tr>
<td>Open-WBO-Inc-Cluster</td>
<td>0.743</td>
</tr>
<tr>
<td>SATLike-c</td>
<td>0.696</td>
</tr>
</tbody>
</table>

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

LinSBPS outperforms Open-WBO-Inc-BMO on 300 seconds by doing a gradual weight-relaxation approach
Incomplete track: Weighted (300 seconds)

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

LinSBPS
Open-WBO-Inc-BMO
maxroster
Open-WBO-Inc-Cluster
SATLike-c
SATLike
Open-WBO-Gluc
Open-WBO-Riss

Score

Instances

LinSBPS
Open-WBO-Inc-BMO
maxroster
Open-WBO-Inc-Cluster
SATLike-c
SATLike
Open-WBO-Gluc
Open-WBO-Riss

Score

0  10  20  30  40  50  60  70  80  90  100  110  120  130  140  150  160  170

Instances

Score

0  0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9  1

Instances

Score

0  10  20  30  40  50  60  70  80  90  100  110  120  130  140  150  160  170

Instances

Score

0  0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9  1

Instances

Score

0  10  20  30  40  50  60  70  80  90  100  110  120  130  140  150  160  170

Instances

Score

0  0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9  1

Instances

Score

0  10  20  30  40  50  60  70  80  90  100  110  120  130  140  150  160  170

Instances

Score
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, Wenzuan 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, João 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/