MaxSAT Evaluation 2017

Carlos Ansótegui
University of Lleida

Fahiem Bacchus
University of Toronto

Matti Järvisalo
University Helsinki

Ruben Martins
Carnegie Mellon University

http://mse17.cs.helsinki.fi/

SAT 2017, August 31, 2017
Outline

- What is new?
- Data
  - Benchmarks
  - Solvers
- Results
  - Complete Tracks
  - Incomplete Tracks
- More information
What is new?

A lot has changed in the MaxSAT Evaluation 2017 (MSE17):

▶ New organization
▶ New rules
▶ New benchmark selection
▶ New evaluation tracks
▶ New ranking for incomplete tracks
▶ New execution environment
New organization

We thank the previous organizers for organizing the MaxSAT Evaluation from 2006 to 2016:

- Josep Argelich Chu, Min Li, Felip Manyà and Jordi Planes

The MSE17 is organized by:

- Carlos Ansótegui (University of Lleida, Spain)
- Fahiem Bacchus (University of Toronto, Canada)
- Matti Järvisalo (University of Helsinki, Finland)
- Ruben Martins (Carnegie Mellon University, US)
New rules

- 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
New benchmark selection

- Complete benchmarks:
  - Benchmark pool: MSE16 benchmarks and new submitted benchmarks
  - Problem:
    - some benchmark sets are much larger than others
  - Solution:
    - Maximum 35 instances per benchmark set
    - Instances selected randomly from the pool of benchmarks

- Incomplete benchmarks:
  - Problem:
    - Complete solvers can solve most benchmarks optimally
  - Solution:
    - Only consider the subset of benchmarks that are not solved optimally under 300 seconds
New evaluation tracks

Evaluation tracks:

- **Unweighted:**
  - Combines the industrial and crafted unweighted and unweighted partial MaxSAT categories from previous MaxSAT evaluations

- **Weighted:**
  - Combines the industrial and crafted weighted and weighted partial MaxSAT categories from previous MaxSAT evaluations

- **Incomplete:**
  - Two special tracks: unweighted and weighted
  - New ranking criterion

- **No-restrictions track:**
  - Portfolio and closed source solvers

MSE 2017 did not include a track for random instances!
New ranking for incomplete tracks

Incomplete ranking:

▶ Before: ranking only considered solvers that got the best solution

▶ Now: we consider how close solvers are to the best solution

▶ 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

$$\sum_i \frac{\text{cost of best solution for } i \text{ found by any solver}}{\text{cost of solution for } i \text{ found by solver}}$$

▶ 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 $[0, 1]$
New execution environment

MSE17 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
New benchmarks

Unweighted (110 new benchmarks):
- extension-enforcement (40)
- min-fill (28)
- gen-hyper-tw (42)

Weighted (700 new benchmarks):
- af-synthesis (40)
- biorepair (30)
- rna-alignment (103)
- css-refactoring (11)
- dalculus (96)
- shiftdesign (30)
- causal-discovery (57)
- metro (30)
- timetabling (30)
- lisbon-wedding (30)
- min-width (222)
- cluster-expansion (21)
MSE17 benchmarks

Complete track:
- Unweighted (880 benchmarks, 97 new)
- Weighted (767 benchmarks, 305 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 (194 benchmarks)
- Weighted (156 benchmarks)
Participating Solvers

1. **LMHS** by Paul Saikko, Tuukka Korhonen, Jeremias Berg and Matti Järvisalo, HIIT, Department of Computer Science University of Helsinki, Finland.

2. **Loandra** by Jeremias Berg, Tuukka Korhonen, and Matti Järvisalo, HIIT, Department of Computer Science University of Helsinki, Finland.

3. **MSUSorting** by Eivind Jahren, Roberto Asín Achá.

4. **MaxHS** by Fahiem Bacchus, University of Toronto, Canada.

5. **MaxRoster** by Takayuki Sugawara, Sugawara Systems, Japan.

6. **Maxino** by Mario Alviano, University of Calabria, Italy.

7. **Open-WBO** by Ruben Martins (Carnegie Mellon University, USA), Miguel Terra-Neves, Saurabh Joshi (IIT-Hyderabad, India), Mikoláš Janota, Vasco Manquinho, Inês Lynce (INESC-ID Portugal).

8. **QMaxSAT** by Naoki Uemura, Aolong Zha, and Miyuki Koshimura, Kyushu University, Japan.
Solvers

Complete track:
- Unweighted: 8 solvers (6 submitters)
- Weighted: 10 solvers (6 submitters)

Incomplete track:
- Unweighted: 4 solvers (4 submitters)
- Weighted: 4 solvers (4 submitters)

Hors Concours solvers:
- Complete unweighted: Open-WBO-MSE16, Z3, CPLEX
- Complete weighted: MaxHS-MSE16, Z3, CPLEX
- Incomplete: CCEHC, Dist, WPM3-in, SAT4J
Solvers

Complete track:

- Unweighted: 8 solvers (6 submitters)
  - Open-WBO (Versions: RES & LSU)
  - MaxHS
  - maxino
  - MSUSorting
  - QMaxSAT (versions: QMaxSAT & uc)
  - LMHS

- Weighted: 10 solvers (6 submitters)
  - Open-WBO (versions: OLL & LSU)
  - MaxHS
  - maxino
  - QMaxSAT (versions: QMaxSAT & uc)
  - LMHS
  - Loandra (versions: S, P & I)
Solvers

Incomplete track:

- **Unweighted**: 4 solvers (4 submitters)
  - maxroster
  - Open-WBO-LSU
  - MaxHS-inc
  - LMHS-inc

- **Weighted**: 4 solvers (4 submitters)
  - maxroster
  - Open-WBO-LSU
  - MaxHS-inc
  - LMHS-inc
Results
Complete track: Unweighted

880 instances

<table>
<thead>
<tr>
<th>Solver</th>
<th>#Solved</th>
<th>Time (Avg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open-WBO-RES</td>
<td>652</td>
<td>129.9</td>
</tr>
<tr>
<td>MaxHS</td>
<td>651</td>
<td>182.61</td>
</tr>
<tr>
<td>maxino</td>
<td>639</td>
<td>99.14</td>
</tr>
<tr>
<td>MSUSorting</td>
<td>622</td>
<td>171.96</td>
</tr>
<tr>
<td>QMaxSATuc</td>
<td>573</td>
<td>165.19</td>
</tr>
</tbody>
</table>

- Best unweighted solvers take advantage of unsatisfiable cores
- How do they compare against last year solvers (Open-WBO-MSE16) and general optimization solvers (Z3, CPLEX)?
Complete track: Unweighted

880 instances

<table>
<thead>
<tr>
<th>Solver</th>
<th>#Solved</th>
<th>Time (Avg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open-WBO-RES</td>
<td>652</td>
<td>129.9</td>
</tr>
<tr>
<td>Open-WBO-MSE16</td>
<td>651</td>
<td>130.61</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Z3</td>
<td>570</td>
<td>187.51</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>CPLEX</td>
<td>392</td>
<td>296.84</td>
</tr>
</tbody>
</table>

- Almost no improvement compared to Open-WBO-MSE16
- Much better than general optimization tools (Z3, CPLEX)
Complete track: Unweighted

Unweighted MaxSAT: Number x of instances solved in y seconds

- Open-WBO-RES
- Open-WBO-MSE16
- MaxHS
- maxino
- MSUSorting
- QMaxSATuc
- Z3
- QMaxSAT
- LMHS
- Open-WBO-LSU
- CPLEX
Complete track: Weighted

767 instances

<table>
<thead>
<tr>
<th>Solver</th>
<th>#Solved</th>
<th>Time (Avg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MaxHS</td>
<td>538</td>
<td>236.46</td>
</tr>
<tr>
<td>QMaxSAT</td>
<td>503</td>
<td>385.18</td>
</tr>
<tr>
<td>QMaxSATuc</td>
<td>499</td>
<td>397.82</td>
</tr>
<tr>
<td>maxino</td>
<td>498</td>
<td>202.1</td>
</tr>
<tr>
<td>Open-WBO-OLL</td>
<td>468</td>
<td>231.88</td>
</tr>
</tbody>
</table>

- MaxHS is much better than the remaining solvers:
  - Uses implicit hitting set approach that combines SAT and IP
- QMaxSAT is a good solver for weighted even though:
  - encodes PB constraints into CNF
  - does not take advantage of unsatisfiable cores
- How do they compare against last year solvers (MaxHS-MSE16) and general optimization solvers (Z3, CPLEX)?
Complete track: Weighted

767 instances

<table>
<thead>
<tr>
<th>Solver</th>
<th>#Solved</th>
<th>Time (Avg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MaxHS</td>
<td>538</td>
<td>236.46</td>
</tr>
<tr>
<td>MaxHS-MSE16</td>
<td>533</td>
<td>225.77</td>
</tr>
<tr>
<td>⋮</td>
<td>⋮</td>
<td>⋮</td>
</tr>
<tr>
<td>Z3</td>
<td>398</td>
<td>260.67</td>
</tr>
<tr>
<td>⋮</td>
<td>⋮</td>
<td>⋮</td>
</tr>
<tr>
<td>CPLEX</td>
<td>348</td>
<td>241.84</td>
</tr>
</tbody>
</table>

- Almost no improvement compared to MaxHS-MSE16
- Much better than general optimization tools (Z3, CPLEX)
Complete track: Weighted

Weighted MaxSAT: Number \( x \) of instances solved in \( y \) seconds

MaxHS
MaxHS-MSE16
QMaxSAT
QMaxSATuc
maxino
Open-WBO-OLL
Loandra-S
Loandra-P
Loandra-I
LMHS
Z3
Open-WBO-LSU
CPLEX
### Increase of time limit to 3600 seconds

What was the effect of increasing the time limit from 1800s to 3600s?

- **Unweighted:**

<table>
<thead>
<tr>
<th>Solver</th>
<th>1800s</th>
<th>3600s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open-WBO-RES</td>
<td>636</td>
<td>652</td>
</tr>
<tr>
<td>MaxHS</td>
<td>636</td>
<td>651</td>
</tr>
<tr>
<td>maxino</td>
<td>631</td>
<td>639</td>
</tr>
<tr>
<td>MSUSorting</td>
<td>606</td>
<td>622</td>
</tr>
<tr>
<td>QMaxSATuc</td>
<td>557</td>
<td>573</td>
</tr>
</tbody>
</table>

- **Weighted:**

<table>
<thead>
<tr>
<th>Solver</th>
<th>1800s</th>
<th>3600s</th>
</tr>
</thead>
<tbody>
<tr>
<td>MaxHS</td>
<td>517</td>
<td>538</td>
</tr>
<tr>
<td>QMaxSAT</td>
<td>470</td>
<td>503</td>
</tr>
<tr>
<td>QMaxSATuc</td>
<td>463</td>
<td>499</td>
</tr>
<tr>
<td>maxino</td>
<td>479</td>
<td>498</td>
</tr>
<tr>
<td>Open-WBO-OLL</td>
<td>446</td>
<td>468</td>
</tr>
</tbody>
</table>

- Usually ~20 more benchmarks solved
- Some solvers benefit more than others
- Ranking would be different in the weighted track!
Incomplete track: Unweighted (60 seconds)

194 instances

<table>
<thead>
<tr>
<th>Solver</th>
<th>Score (avg)</th>
<th>#Solutions</th>
<th>#Best Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open-WBO-LSU</td>
<td>0.732</td>
<td>175</td>
<td>56</td>
</tr>
<tr>
<td>MaxHS-inc</td>
<td>0.662</td>
<td>177</td>
<td>17</td>
</tr>
<tr>
<td>maxroster</td>
<td>0.643</td>
<td>141</td>
<td>75</td>
</tr>
<tr>
<td>WPM3-in †</td>
<td>0.625</td>
<td>193</td>
<td>25</td>
</tr>
<tr>
<td>SAT4J †</td>
<td>0.585</td>
<td>161</td>
<td>15</td>
</tr>
<tr>
<td>LMHS-inc</td>
<td>0.561</td>
<td>157</td>
<td>21</td>
</tr>
<tr>
<td>Dist †</td>
<td>0.526</td>
<td>147</td>
<td>48</td>
</tr>
<tr>
<td>CCEHC †</td>
<td>0.526</td>
<td>124</td>
<td>65</td>
</tr>
</tbody>
</table>

† Hors concours solver

- Open-WBO-LSU is not a good complete solver but can quickly find high quality intermediate solutions for unweighted MaxSAT
- Score-based ranking favors solvers that find solutions
- Prior ranking based on finding the best solution is very different!
Incomplete track: Unweighted (300 seconds)

194 instances

<table>
<thead>
<tr>
<th>Solver</th>
<th>Score (avg)</th>
<th>#Solution</th>
<th>#Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>maxroster</td>
<td>0.846</td>
<td>179</td>
<td>116</td>
</tr>
<tr>
<td>Open-WBO-LSU</td>
<td>0.694</td>
<td>175</td>
<td>43</td>
</tr>
<tr>
<td>MaxHS-inc</td>
<td>0.670</td>
<td>182</td>
<td>33</td>
</tr>
<tr>
<td>SAT4J†</td>
<td>0.593</td>
<td>175</td>
<td>16</td>
</tr>
<tr>
<td>CCEHC†</td>
<td>0.580</td>
<td>149</td>
<td>57</td>
</tr>
<tr>
<td>LMHS-inc</td>
<td>0.570</td>
<td>176</td>
<td>20</td>
</tr>
<tr>
<td>WPM3-in†</td>
<td>0.552</td>
<td>193</td>
<td>20</td>
</tr>
<tr>
<td>Dist†</td>
<td>0.522</td>
<td>151</td>
<td>48</td>
</tr>
</tbody>
</table>

† Hors concours solver

- maxroster is much better with 300 seconds:
  - It outperforms the other solvers in both score and best solutions!
- Stochastic solver CCEHC can often find the best solution
Incomplete track: Weighted (60 seconds)

156 instances

<table>
<thead>
<tr>
<th>Solver</th>
<th>Score (avg)</th>
<th>#Solution</th>
<th>#Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>maxroster</td>
<td>0.800</td>
<td>147</td>
<td>65</td>
</tr>
<tr>
<td>WPM3-in†</td>
<td>0.758</td>
<td>151</td>
<td>24</td>
</tr>
<tr>
<td>SAT4J†</td>
<td>0.751</td>
<td>146</td>
<td>16</td>
</tr>
<tr>
<td>LMHS-inc</td>
<td>0.711</td>
<td>146</td>
<td>10</td>
</tr>
<tr>
<td>Open-WBO-LSU</td>
<td>0.677</td>
<td>141</td>
<td>37</td>
</tr>
<tr>
<td>MaxHS-inc</td>
<td>0.669</td>
<td>141</td>
<td>17</td>
</tr>
<tr>
<td>Dist†</td>
<td>0.509</td>
<td>98</td>
<td>27</td>
</tr>
<tr>
<td>CCEHC†</td>
<td>0.473</td>
<td>90</td>
<td>27</td>
</tr>
</tbody>
</table>

† Hors concours solver

- maxroster clearly outperforms other solvers
- Stochastic solvers often find the best solution
Incomplete track: Weighted (300 seconds)

156 instances

<table>
<thead>
<tr>
<th>Solver</th>
<th>Score (avg)</th>
<th>#Solution</th>
<th>#Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>maxroster</td>
<td>0.834</td>
<td>149</td>
<td>78</td>
</tr>
<tr>
<td>WPM3-in†</td>
<td>0.767</td>
<td>152</td>
<td>19</td>
</tr>
<tr>
<td>SAT4J†</td>
<td>0.766</td>
<td>151</td>
<td>8</td>
</tr>
<tr>
<td>MaxHS-inc</td>
<td>0.760</td>
<td>148</td>
<td>26</td>
</tr>
<tr>
<td>LMHS-inc</td>
<td>0.740</td>
<td>145</td>
<td>6</td>
</tr>
<tr>
<td>Dist†</td>
<td>0.523</td>
<td>104</td>
<td>22</td>
</tr>
<tr>
<td>CCEHC†</td>
<td>0.519</td>
<td>103</td>
<td>20</td>
</tr>
<tr>
<td>Open-WBO-LSU</td>
<td>0.496</td>
<td>92</td>
<td>34</td>
</tr>
</tbody>
</table>

† Hors concours solver

- maxroster clearly outperforms other solvers
- Score ranking differs substantially from best ranking
- Open-WBO-LSU cannot find many solutions because it was killed for reaching the memory limit before outputting the best solution
No restrictions track

- Only one submission

- Solver “aurora borealis”:
  - 1QBit company
  - Incomplete solver
  - Only supports unweighted MaxSAT without hard clauses
  - Did not compete since unweighted MaxSAT track contains hard clauses
MaxSAT Evaluation 2017 webpage

http://mse17.cs.helsinki.fi/

- 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
- Partial description of the benchmarks
- Benchmarks and log files are available for download
- SQLite database with all results
Achievements and Failures

Achievements:

- We now have several available open source MaxSAT solvers
- All MaxSAT solvers have a description using IEEE Proceedings style
- Benchmark submission was high (particularly for weighted problems)
- Benchmark selection is more balanced between benchmark domains than before

Failures:

- Still missing descriptions for several benchmark domains
- Lower solver participation in the complete track
- Incomplete track did not have many solvers since most incomplete solvers target random instances
Thanks

Thanks to the people that contributed solvers and benchmarks:

Thanks to StarExec for allowing us to use their cluster:

https://www.starexec.org/