

# 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)
- ▶ dcalculus (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

Solver	#Solved	Time (Avg)
Open-WBO-RES	652	129.9
MaxHS	651	182.61
maxino	639	99.14
MSUSorting	622	171.96
QMaxSATuc	573	165.19

- ▶ 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

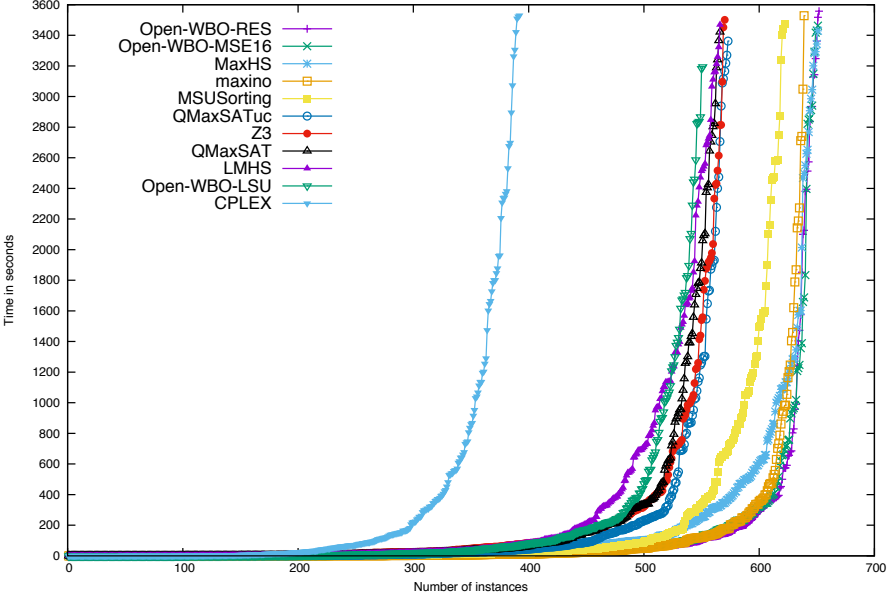
880 instances

Solver	#Solved	Time (Avg)
Open-WBO-RES	652	129.9
Open-WBO-MSE16	651	130.61
...	...	...
Z3	570	187.51
...	...	...
CPLEX	392	296.84

- ▶ 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



# Complete track: Weighted

767 instances

Solver	#Solved	Time (Avg)
MaxHS	538	236.46
QMaxSAT	503	385.18
QMaxSAT <sub>uc</sub>	499	397.82
maxino	498	202.1
Open-WBO-OLL	468	231.88

- ▶ 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

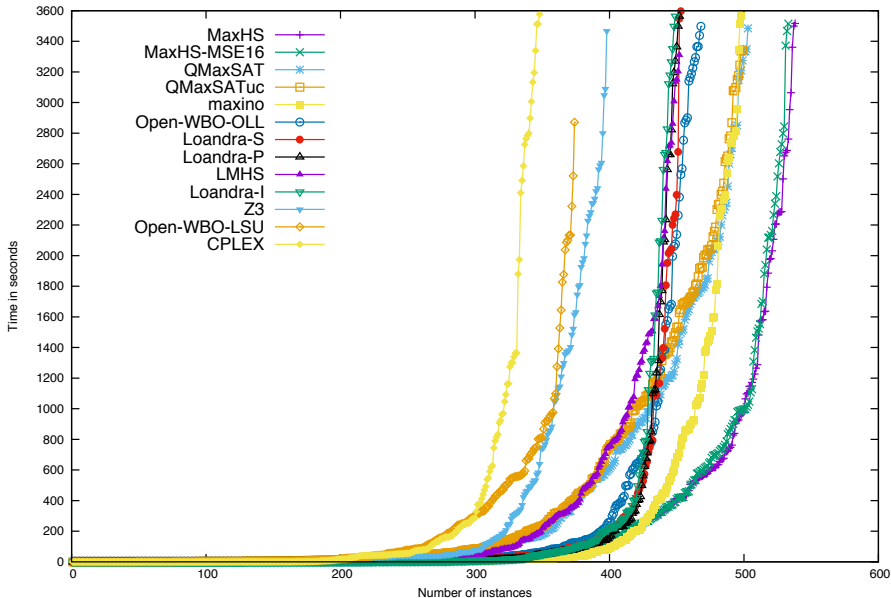
767 instances

Solver	#Solved	Time (Avg)
MaxHS	538	236.46
MaxHS-MSE16	533	225.77
...	...	...
Z3	398	260.67
...	...	...
CPLEX	348	241.84

- ▶ 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



# Increase of time limit to 3600 seconds

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

► Unweighted:

Solver	1800s	3600s
Open-WBO-RES	636	652
MaxHS	636	651
maxino	631	639
MSUSorting	606	622
QMaxSATuc	557	573

► Weighted:

Solver	1800s	3600s
MaxHS	517	538
QMaxSAT	470	503
QMaxSATuc	463	499
maxino	479	498
Open-WBO-OLL	446	468

- 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

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

<sup>†</sup> 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

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

<sup>†</sup> 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

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

<sup>†</sup> Hors concours solver

- ▶ maxroster clearly outperforms other solvers
- ▶ Stochastic solvers often find the best solution

# Incomplete track: Weighted (300 seconds)

156 instances

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

<sup>†</sup> 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

# Webpage

## 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/>

