Graph Neural Networks for Dynamic Scheduling of SMT Solvers

Jan Hůla, David Mojžíšek, Mikoláš Janota

CZECH INSTITUTE OF INFORMATICS ROBOTICS AND CYBERNETICS CTU IN PRAGUE

Satisfiability Modulo Theories

- Deals with a decision problem for logical formulas written in FOL with dedicated theory.
- Various solvers have been developed for different theories.

```
(set-info :smt-lib-version 2.6)
(set-logic QF_NRA)
(set-info :category "crafted")
(set-info :status sat)
(declare-fun a () Real)
(assert (and (>= a 3)
  (not (>= (* a 2) 3))))
(check-sat)
(exit)
```

Solver portfolios

- Popularized by SATzilla (E. Nudelman et al., 2004)
- Empirical hardness models - models trained to predict runtime of an algorithm on a given input.
- Because of the uncertainty in the predictions, schedule of solvers may be more effective then solver selection.

Comparing solvers

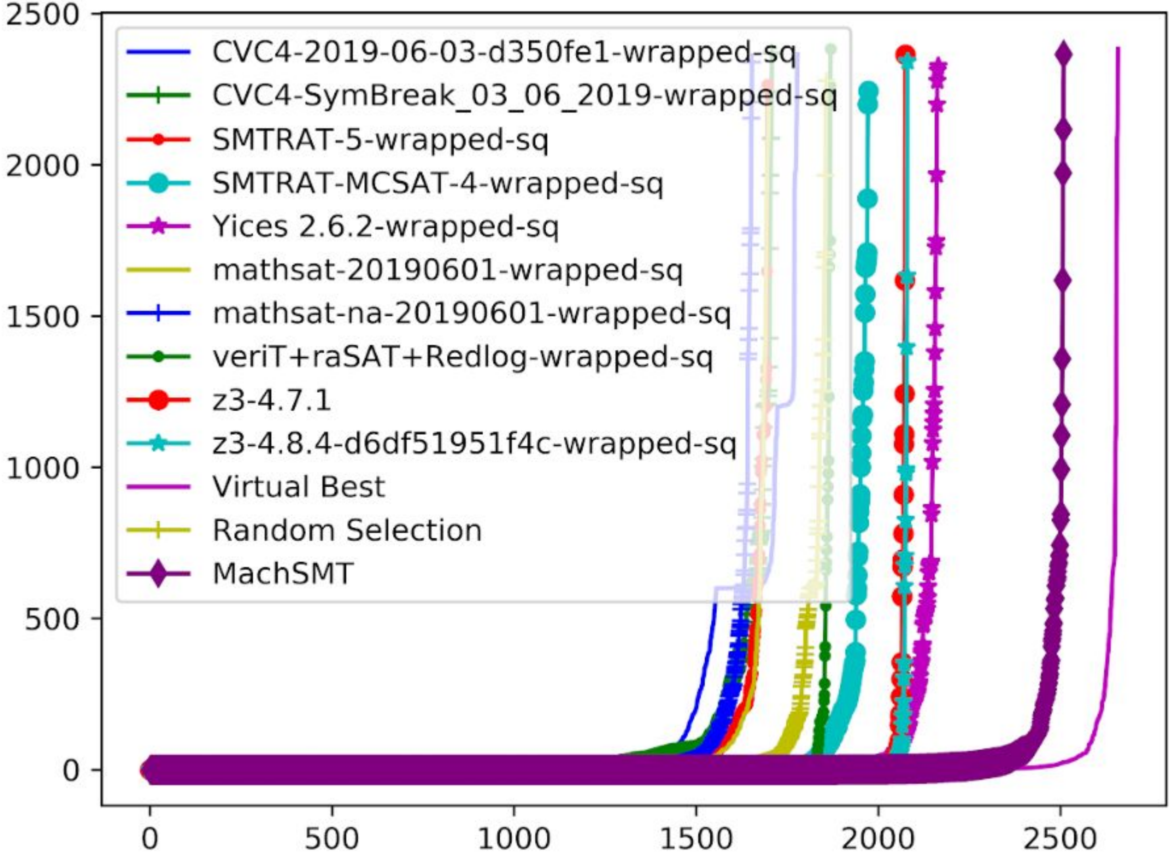
- SMT-COMP - competition of solvers on in different benchmark sets
- PAR2 - penalized average runtime (with penalization 2*timeout)
- Virtual best solver - gives an upper bound; picks the best solver on every instance.

Related work

- **MachSMT**: A Machine Learning-based Algorithm Selector for SMT Solvers (Scott et al., 2020)
- BOW representation of formulas
- Boosted trees as empirical hardness model

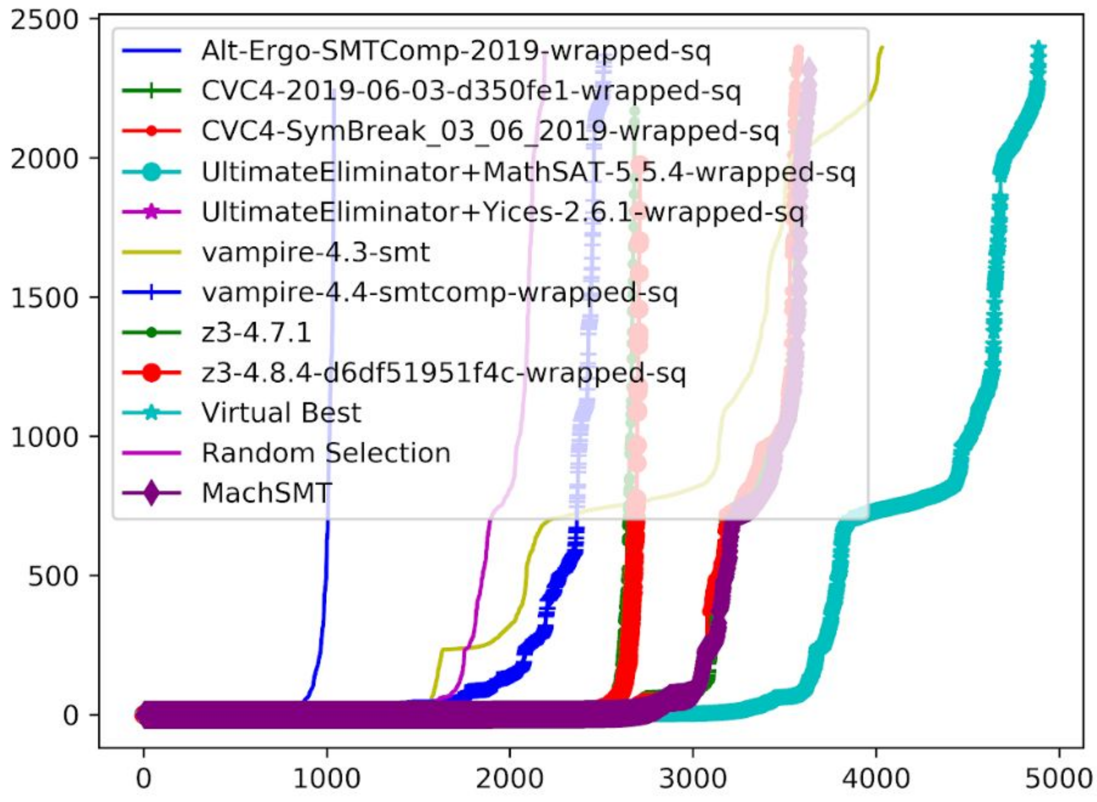
MachSMT - results

QFNRA



MachSMT - results

UFNIA



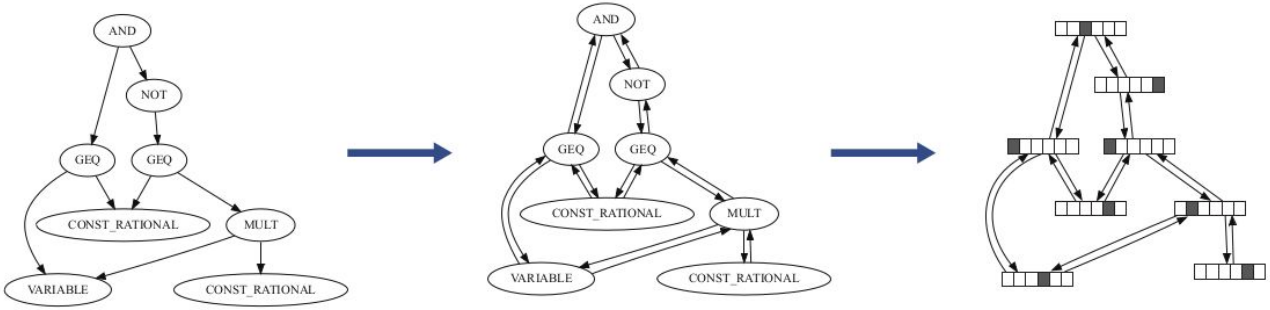
Selection of benchmark sets

- Requirements:
  - enough examples
  - a gap between the best solver/MachSMT and the virtual best solver

Benchmark name	# of problems	# of solvers	Timeout (s)
QF_NRA	2654	9	2400
UFNIA	5659	7	2400
UFNIA-CONF	5659	23	60
TPTP	4741	25	300

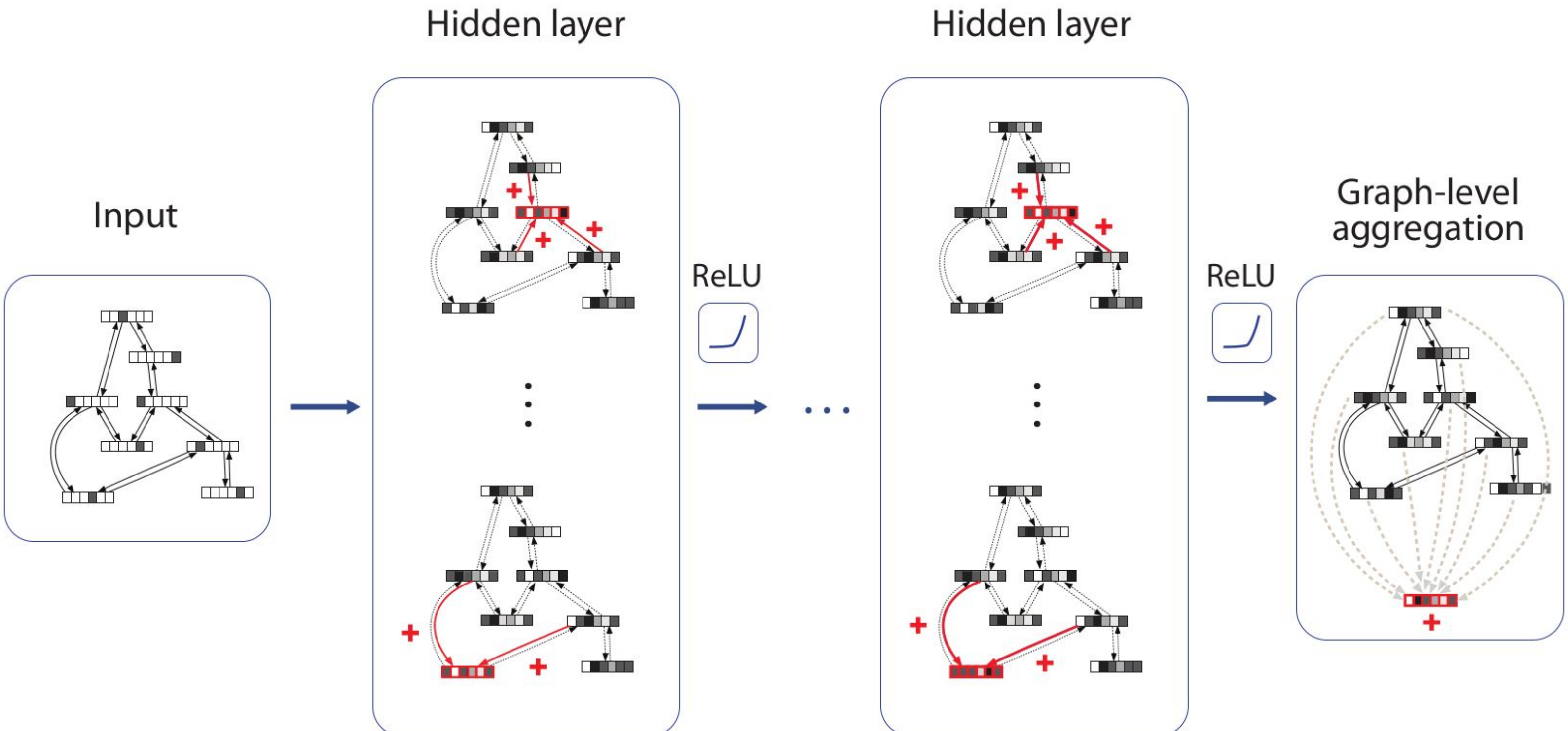
Encoding of SMT formulas

```
(set-info :smt-lib-version 2.6)
(set-logic QF_NRA)
(set-info :category "crafted")
(set-info :status sat)
(declare-fun a () Real)
(assert (and (>= a 3)
  (not (>= (* a 2) 3))))
(check-sat)
(exit)
```



Our model

- GCN (Kipf et al., 2016), 6 layers



Results for solver selection

- GNN beats MachSMT 👍
- Random schedule beats both 👎
- Most problems are solved in few seconds at least by some solver.

BOW single	PAR-2 impr.	117.10%
	solved	2343
GNN single	PAR-2 impr.	231.19%
	solved	2403
Random schedule	PAR-2 impr.	269.33%
	solved	2494

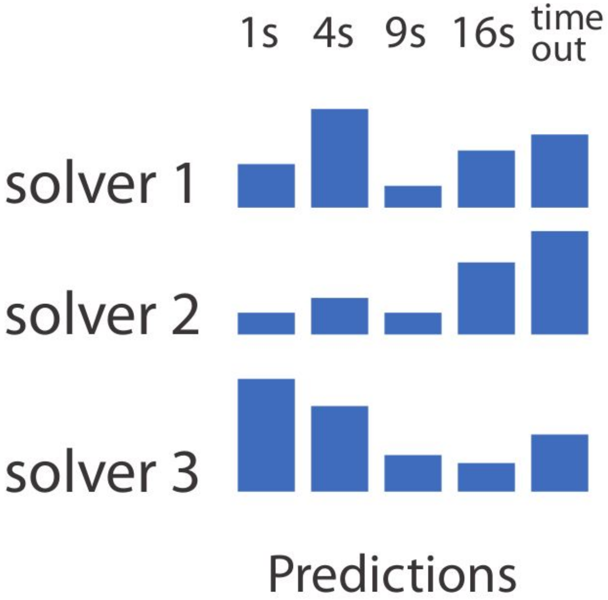
Solver scheduling

- Basic schedule: order solvers according to the predictions of the runtime

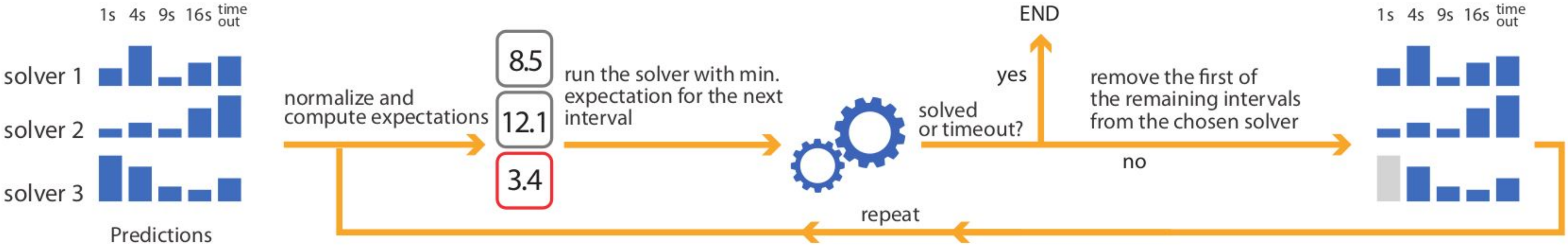
BOW single	PAR-2 impr.	117.10%
	solved	2343
GNN single	PAR-2 impr.	231.19%
	solved	2403
Random schedule	PAR-2 impr.	269.33%
	solved	2494
Solver ordering	PAR-2 impr.	913.05%
	solved	2494

Prediction of solving time distribution

- We split the available runtime to multiple intervals and train the GNN to classify in which interval the problem will be solved by a given solver.
- The predictions are “probabilistic”.
- Can compute approximate expected runtime



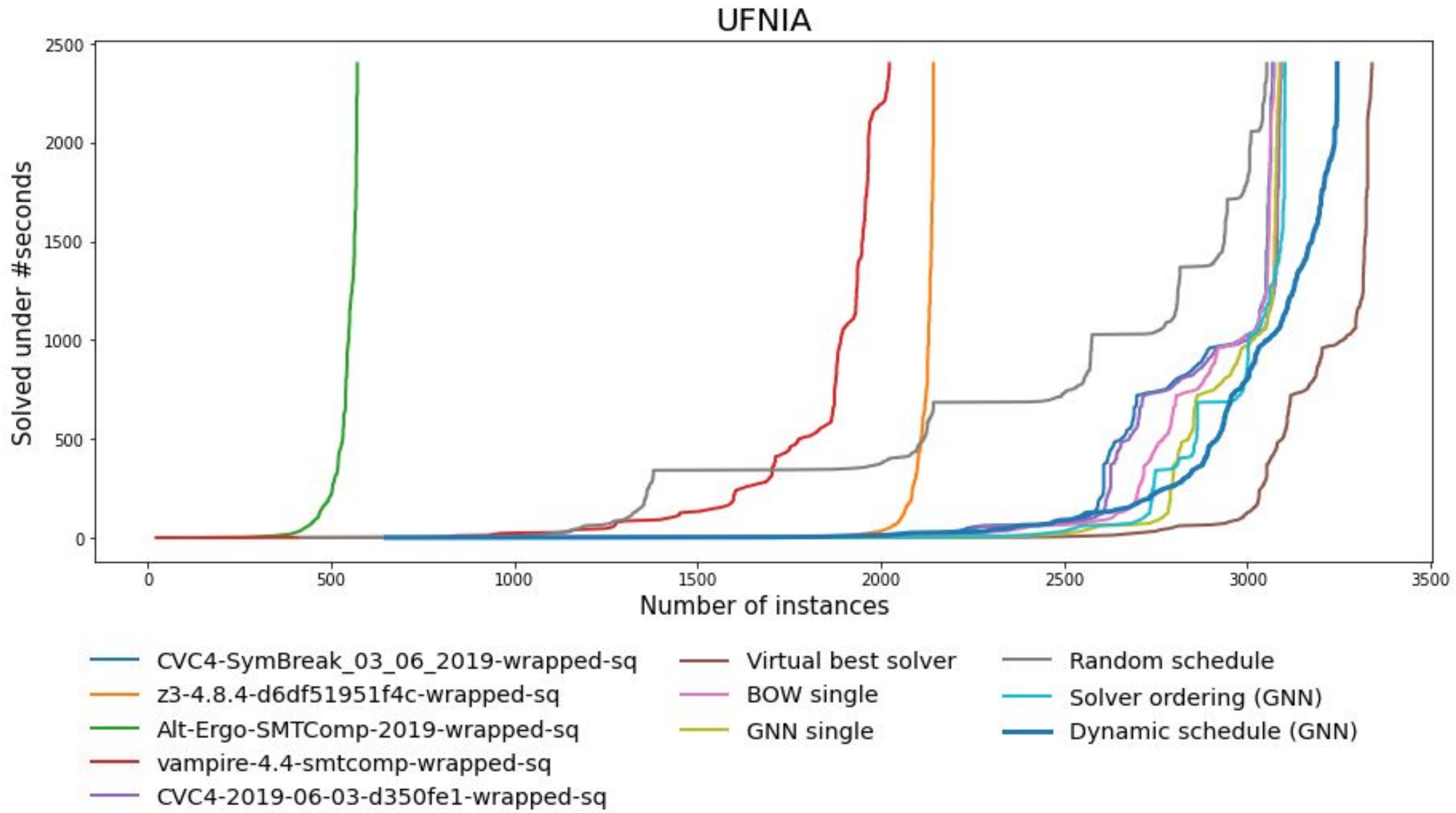
Scheduling solvers using expected runtime



Interval lengths

- “In empirical benchmark sets, the probability that a problem will be solved decreases exponentially with time” (Pimpalkhare, 2021)
- The exponent is different for different solvers
- We estimate these exponents from a training set and split than available time to chunks of exponentially increasing lengths.

Results


UFNIA
Solved under #seconds
0
500
1000
1500
2000
2500
0
500
1000
1500
2000
2500
3000
3500
Number of instances
CVC4-SymBreak_03_06_2019-wrapped-sq
z3-4.8.4-d6df51951f4c-wrapped-sq
Alt-Ergo-SMTComp-2019-wrapped-sq
vampire-4.4-smtcomp-wrapped-sq
CVC4-2019-06-03-d350fe1-wrapped-sq
Virtual best solver
BOW single
GNN single
Random schedule
Solver ordering (GNN)
Dynamic schedule (GNN)

Results

Benchmark		QF-NRA	UFNIA	UFNIA-CONF	TPTP
Best Solver	solver	Z3	CVC4	-	Vampire
	solved	2120	3093	2494	3204
VBS	solved	2516	3339	3118	3564
BOW single	PAR-2 impr.	117.10%	-0.64%	0.32%	13.56%
	solved	2343	3074	2586	3315
GNN single	PAR-2 impr.	231.19%	1.8%	56.73%	18.23%
	solved	2403	3085	2644	3320
Random schedule	PAR-2 impr.	269.33%	-21.62%	69.59%	-15.05%
	solved	2494	3053	2812	3266
Solver ordering	PAR-2 impr.	913.05%	8.30%	88.59%	6.84%
	solved	2494	3053	2812	3266
Dynamic schedule	PAR-2 impr.	**1162.87%**	**76.91%**	**113.54%**	**51.92%**
	solved	**2498**	**3245**	**2891**	**3388**

Final remarks

- Exploiting dataset correlations vs OOD generalization.
- Future work: adapting for distribution shift.