Automatic Configuration of Benchmark Sets for Classical Planning

Álvaro Torralba,[1] Jendrik Seipp,[2] Silvan Sievers[2]
[1]Aalborg University, Denmark
[2]University of Basel, Switzerland

October 21, 2020

Outline

The Cycle of Life (in Planning Research)

Everything you Always Wanted to Know About Planning
(But Were Afraid to Ask) — (Jörg Hoffmann, 2011)

The Cycle of Life (in Planning Research)

Everything you Always Wanted to Know About Planning (But Were Afraid to Ask) — (Jörg Hoffmann, 2011)

Is, then, the life of a researcher in heuristic search planning characterized by the following pseudo-code?

while (not retired) **do**
    think up some new heuristic $h^{foo\text{-}bar}$
    run it on the benchmarks
endwhile

Fig. 3. The life of a planning researcher?

The Cycle of Life (in Planning Research)

Everything you Always Wanted to Know About Planning (But Were Afraid to Ask) — (Jörg Hoffmann, 2011)

Is, then, the life of a researcher in heuristic search planning characterized by the following pseudo-code?

while (not retired) **do**
    think up some new heuristic $h^{foo\text{-}bar}$
    run it on the benchmarks
endwhile

Fig. 3. The life of a planning researcher?

The answer to that one is "NO!". Far beyond just improving performance on benchmarks, the *understanding* of heuristics is where heuristic search planning really turns into a natural science. Dramatic progress has been made, in that science, during the last years. For example, Bonet and Geffner [3] proved

Empirical Evaluation – Examples from HSDIP'20

	Coverage			Time Score		
	offline	on-sushi	on-oshibi	offline	on-sushi	on-oshibi
agricola (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
airport (50)	**34**	24	**34**	2.5	18.5	**26.1**
barman (34)	4	0	**4**	0.3	0.0	**2.0**
blocks (35)	**28**	20	**28**	2.2	19.5	**29.1**
childsnack (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
data-network (20)	**14**	11	**14**	1.1	7.2	**12.1**
depot (22)	13	6	**13**	1.0	3.3	9.6
driverlog (20)	**15**	7	**15**	1.1	4.3	**10.5**
elevators (30)	**44**	12	**44**	3.4	2.6	25.9
floortile (40)	6	0	6	0.3	0.0	**0.8**
freecell (80)	**68**	30	**68**	4.8	10.9	35.8
ged (20)	**19**	7	**19**	1.5	4.7	12.1
grid (5)	3	1	3	0.2	1.0	2.2
gripper (20)	8	6	**8**	0.6	4.8	**7.0**
hiking (20)	14	8	**15**	1.0	5.1	**10.9**
logistics (63)	**39**	19	**39**	2.8	11.8	25.2
miconic (150)	**144**	133	**144**	11.2	80.6	131.6
movie (30)	**30**	**30**	**30**	2.4	**42.7**	42.4
mprime (35)	**29**	24	**29**	2.3	19.1	**26.7**
mystery (30)	**19**	15	**19**	1.5	12.6	**17.8**
nomystery (20)	**20**	12	**20**	1.5	8.0	15.3
openstacks (100)	**53**	21	**53**	3.8	12.0	29.8
organic (20)	7	7	7	0.5	6.0	**6.1**
organic-split (20)	**10**	6	**10**	0.7	1.9	**4.2**
parcprinter (50)	**38**	34	**38**	2.9	28.2	34.1
parking (40)	**13**	1	**13**	0.9	0.1	**5.4**
pathways (30)	5	4	5	0.3	5.1	5.8
pegsol (50)	**48**	42	**48**	3.7	22.8	**35.6**
petri-net (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
pipes-nt (50)	**25**	14	**25**	1.8	10.4	**19.0**
pipes-t (50)	**18**	8	**18**	1.3	5.1	**12.1**
psr-small (50)	**50**	49	**50**	3.9	48.2	**54.6**
rovers (40)	**8**	7	**8**	0.6	6.6	**8.1**
satellite (36)	7	6	7	0.5	5.3	**7.2**
scanalyzer (50)	**35**	7	33	**35**	2.7	5.7
snake (20)	12	6	**12**	0.9	2.5	**7.6**
sokoban (50)	**50**	33	**50**	**50**	3.8	19.6
spider (20)	**15**	7	**15**	1.1	2.9	**8.6**
storage (30)	**16**	14	**16**	**16**	1.2	12.5
termes (20)	12	0	**12**	12	0.8	0.0
tetris (17)	**11**	7	10	**11**	0.8	1.3
tidybot (40)	**25**	18	24	**25**	1.8	5.8
tpp (30)	**8**	5	**8**	**8**	0.6	8.0
transport (70)	34	20	**36**	**36**	2.6	10.4
trucks (30)	**13**	9	**13**	**13**	0.9	5.3
visitall (40)	30	**33**	30	30	2.3	27.8
woodwork (50)	**49**	38	**49**	**49**	3.8	24.3
zenotravel (20)	**13**	7	**13**	**13**	1.0	4.9
Sum (1827)	1156	766	1155	**1159**	86.8	539.8
						900.4
						908.7

Empirical Evaluation – Examples from HSDIP'20

	Coverage			Time Score		
	offline	on-undef	on-solv	offline	on-undef	on-solv
agricola (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
airport (50)	**34**	34	**34**	2.5	18.5	**26.1**
barman (34)	4	0	**4**	0.3	0.0	**2.0**
blocks (35)	**28**	**20**	**28**	2.2	19.5	**29.1**
childsnack (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
data-network (20)	**14**	**11**	**14**	1.1	7.2	**12.1**
depot (22)	**13**	6	**13**	1.0	3.3	9.6
driverlog (20)	**15**	**7**	**15**	1.1	4.3	**10.7**
elevators (30)	**44**	**12**	**44**	**3.4**	2.6	**25.9**
floortile (40)	6	0	**6**	0.3	0.0	**0.8**
freecell (80)	**68**	**30**	**68**	**4.8**	10.9	**35.8**
ged (20)	**19**	**7**	**19**	**1.5**	4.7	**12.1**
grid (5)	**3**	1	**3**	0.2	1.0	**2.3**
gripper (20)	**8**	**6**	**8**	**0.6**	4.8	**7.0**
hiking (20)	14	8	**15**	1.0	5.1	**10.9**
logistics (63)	**39**	**19**	**39**	**3.9**	11.8	**25.2**
miconic (150)	**144**	**133**	**144**	**144**	11.2	**80.6**
movie (30)	**30**	**30**	**30**	**30**	2.4	**42.7**
mprime (35)	**29**	**24**	**29**	**29**	2.3	**19.1**
mystery (30)	**19**	**15**	**19**	**19**	1.5	**12.6**
nomystery (20)	**20**	**12**	**20**	**20**	1.5	**8.0**
openstacks (100)	**53**	**21**	**53**	**53**	3.8	**12.0**
organic (20)	7	**7**	**7**	0.7	0.5	**6.0**
organic-split (20)	**10**	**6**	**10**	**10**	0.7	**1.9**
pare-printer (50)	**38**	**14**	**38**	**38**	2.9	**28.2**
parking (40)	**13**	**1**	**13**	**13**	0.9	**0.1**
pathways (30)	5	4	**5**	**5**	0.3	**5.1**
pegsol (50)	**48**	**42**	**48**	**48**	3.7	**22.8**
petri-net (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
pipes-nt (50)	**25**	14	**25**	**25**	1.8	**10.4**
pipes-t (50)	**18**	**8**	**18**	**18**	1.3	**5.1**
psr-small (50)	**50**	**49**	**50**	**50**	3.9	**48.2**
rovers (40)	**8**	**7**	**8**	**8**	0.6	**6.6**
satellite (36)	7	6	**7**	**7**	0.5	**5.3**
scanalyzer (50)	**35**	**7**	**33**	**35**	2.7	**5.7**
snake (20)	**12**	6	**12**	**12**	0.9	**2.5**
sokoban (50)	**50**	**13**	**50**	**50**	3.8	**19.6**
spider (20)	**15**	**7**	**15**	**15**	1.1	**2.9**
storage (30)	**16**	**14**	**16**	**16**	1.2	**12.5**
termes (20)	**12**	0	**12**	**12**	0.8	**0.0**
tetris (17)	**11**	**7**	**10**	**11**	0.8	**1.3**
tidybot (40)	**25**	**18**	**24**	**25**	1.8	**5.8**
tpp (30)	**8**	**7**	**8**	**8**	0.6	**8.0**
transport (50)	34	**20**	**36**	**36**	2.6	**10.4**
trucks (30)	**13**	**9**	**13**	**13**	0.9	**5.3**
visitall (40)	30	**33**	30	30	2.3	**27.8**
woodwork (50)	**49**	**31**	**49**	**49**	3.8	**24.3**
zenotravel (20)	**13**	**7**	**13**	**13**	1.0	**4.9**
Sum (1827)	1156	766	1155	**1159**	86.8	539.8

	Total	p_1	p_2	p_3	$\|\text{SOCG}_1\|$	τ_1^1	τ_2^1	τ_1^2	τ_2^2
agricola	0	0	0	**16**	1	13	12	11	9
airport	23	35	**36**	33	1	21	13	14	17
barman	0	0	12	**20**	0	0	0	0	0
blocks	18	**35**	**35**	**35**	21	17	17	**35**	**35**
childsnack	0	0	1	0	0	0	0	0	0
data-network	0	1	1	5	2	0	0	0	0
depot	4	14	**18**	**18**	6	5	5	15	13
driverlog	7	**19**	18	18	8	12	13	15	14
elevators	0	5	**20**	7	0	0	0	0	0
floortile	0	0	2	1	0	0	0	0	0
freecell	20	46	79	**80**	18	**80**	23	61	57
ged	0	**20**	**20**	**20**	0	**20**	0	0	0
grid	1	3	4	**5**	0	3	0	3	4
gripper	**8**	**20**	**20**	**20**	**8**	**20**	10	**20**	**20**
hiking	2	3	**20**	**20**	2	7	5	3	3
logistics	2	7	**29**	15	2	2	3	6	5
maintenance	0	**14**	11	14	0	0	0	0	0
miconic	55	**150**	**150**	**150**	71	**150**	**146**	**150**	**150**
movie	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**
mprime	20	21	**32**	22	5	18	13	18	17
mystery	15	15	**17**	14	9	6	12	10	9
openstacks	0	0	2	**20**	0	1	8	2	5
organic-synthesis	**3**	**3**	2	**3**	**3**	**3**	**3**	**3**	2
parepristar	0	12	**20**	18	0	0	1	0	0
parking	0	0	7	0	0	0	0	0	0
pathways	4	5	**10**	8	4	4	4	4	4
pegsol	17	**20**	**20**	**20**	18	**20**	**20**	**20**	**20**
pipesworld	12	22	23	**27**	12	13	15	20	17
psr	1	49	**50**	**50**	**50**	**50**	**50**	**50**	**50**
rovers	6	21	**28**	25	6	14	16	16	16
satellite	6	15	**27**	12	7	15	8	8	9
scanalyzer	4	**20**	18	**20**	5	18	11	**20**	18
snake	3	4	5	7	4	3	3	7	6
sokoban	6	13	19	10	8	11	12	10	9
spider	1	12	9	15	0	5	7	5	9
storage	14	18	**19**	**19**	17	15	**19**	18	**19**
termes	0	10	14	**15**	3	4	2	7	3
tetris	0	**20**	9	**20**	2	11	19	2	18
thoughtful	5	5	8	**14**	5	12	5	5	5
tidybot	3	19	16	**20**	0	1	7	8	2
tpp	6	13	23	**29**	6	15	10	14	15
transport	0	5	9	**16**	0	0	0	0	0
trucks	6	9	**15**	9	6	6	7	7	7
visitall	0	**20**	9	**20**	0	0	9	0	0
woodworking	1	1	2	**4**	1	1	1	1	1
zenotravel	8	**20**	**20**	**20**	7	9	9	13	14
SUM	359	775	933	**965**	346	451	358	431	432
$>h_{\text{sa}}$					4	6	4	6	7
$>h_{\text{ff}}$					2	5	5	4	3
$>h_{\text{lm}}$					0	2	1	0	0

Empirical Evaluation – Examples from HSDIP'20

	Coverage			Time Score		
	offline	on-node	on-sat4e	offline	on-node	on-sat4e
agricola (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
airport (50)	34	24	34	2.5	18.5	**26.1**
barman (34)	4	0	4	0.3	0.0	**2.0**
blocks (35)	**28**	**28**	**28**	2.2	19.5	**29.1**
childsnack (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
data-network (20)	14	11	14	1.1	7.2	**12.1**
depot (22)	13	6	13	1.0	3.3	9.6
diverbeg (20)	15	7	15	1.1	4.3	**10.7**
elevators (50)	44	12	44	3.4	2.6	25.9
floortile (40)	6	0	6	0.3	0.0	**0.8**
freecell (80)	68	30	68	4.8	10.9	35.8
ged (20)	19	7	19	1.5	4.7	12.1
grid (5)	3	1	3	0.2	1.0	2.2
gripper (20)	8	6	8	0.6	4.8	**7.0**
hiking (20)	14	8	**15**	1.0	5.1	**10.9**
logistics (63)	39	19	39	2.8	11.8	25.2
miconic (150)	**144**	**133**	**144**	11.2	80.6	131.6
movie (30)	**30**	**30**	**30**	2.4	**42.7**	42.4
mprime (35)	29	24	29	2.3	19.1	26.5
mystery (30)	19	**15**	19	1.5	12.6	**17.8**
nomystery (20)	20	12	20	1.5	8.0	15.3
openstacks (100)	53	21	**53**	3.8	12.0	29.8
organic (20)	7	7	7	0.5	6.0	**6.1**
organic-split (20)	10	6	10	0.7	1.9	**4.2**
parcprinter (30)	**35**	**34**	**35**	2.9	28.2	34.1
parking (40)	13	1	13	0.9	0.1	**5.4**
pathways (30)	5	4	5	0.3	3.1	**5.5**
pegsol (36)	**48**	**42**	**48**	**48**	3.7	**22.8**
petri-net (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
pipes-nt (50)	**25**	**14**	**25**	1.8	10.4	**19.0**
pipes-t (50)	**18**	**8**	**18**	1.3	5.1	**12.1**
psr-small (50)	**50**	**49**	**50**	3.9	**48.2**	**54.6**
rovers (40)	8	7	8	0.6	6.6	**8.1**
satellite (36)	7	6	7	0.5	5.5	7.3
scanalyzer (50)	**35**	7	**35**	2.7	5.7	19.6
snake (20)	12	6	**12**	0.9	2.5	**7.6**
sokoban (50)	**50**	**33**	**50**	**50**	3.8	19.6
spider (20)	**15**	7	**15**	1.1	2.9	**8.6**
storage (30)	**16**	**14**	**16**	1.2	12.5	**17.1**
termes (20)	12	0	**12**	0.8	0.0	**3.2**
tetris (17)	**11**	**3**	**11**	0.8	1.3	5.4
tidybot (40)	**18**	**16**	**18**	1.8	5.8	**15.1**
tpp (30)	**8**	**7**	**8**	0.6	8.0	**8.9**
transport (70)	34	**20**	**36**	**36**	2.6	**22.4**
trucks (30)	**13**	**9**	**13**	**13**	0.9	5.3
visitall (40)	30	**33**	30	30	2.3	**27.8**
woodworking (50)	**49**	**58**	**49**	**49**	3.8	**24.3**
zenotravel (20)	**13**	7	**13**	**13**	1.0	**4.9**
Sum (1672)	1156	766	1155	**1159**	86.6	**539.9**

	Total	IC	MT	Min	SINGL (Min)	C1	C2	C3	C4
agricola	0	0	0	**16**	1	13	12	11	9
airport	23	35	**36**	33	1	21	13	14	17
barman	0	0	12	**20**	0	0	0	0	0
blocks	18	**35**	**35**	**35**	21	33	27	**35**	**35**
childsnack	0	0	1	0	0	0	0	0	0
data-network	0	1	5	2	0	0	0	0	0
depot	4	14	**18**	**18**	6	5	5	15	13
diverbeg	2	**19**	**19**	18	8	12	15	15	14
elevators	0	5	**20**	7	0	0	0	0	0
floortile	0	0	2	1	0	0	0	0	0
freecell	20	46	79	**80**	18	**80**	23	61	57
ged	0	**20**	**20**	**20**	0	**20**	0	0	0
grid	1	3	4	5	0	3	0	3	4
gripper	8	**20**	**20**	**20**	8	**20**	10	**20**	**20**
hiking	2	**20**	**20**	**20**	2	7	5	3	3
logistics	2	7	**29**	15	2	2	3	6	5
maintenance	0	14	14	11	0	0	0	0	0
miconic	55	**150**	**150**	**150**	71	**150**	146	**150**	**150**
movie	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**
mprime	20	21	**32**	22	5	10	13	18	17
mystery	15	15	**17**	14	9	6	12	10	9
openstacks	0	0	2	**20**	0	1	8	2	5
organic-synthesis	**3**	**3**	2	**3**	**3**	**3**	**3**	**3**	2
parcprinter	0	12	**20**	18	0	0	1	0	0
parking	0	0	7	0	0	0	0	0	0
pathways	4	5	**10**	8	4	4	4	4	4
pegsol	17	**20**	**20**	**20**	18	**20**	**20**	**20**	**20**
pipesworld	12	22	23	**27**	12	13	15	20	**27**
psr	49	**50**	**50**	**50**	**50**	**50**	**50**	**50**	**50**
rovers	6	21	**26**	25	6	14	16	16	16
satellite	6	15	**20**	12	6	15	8	9	9
scanalyzer	4	**20**	18	**20**	5	18	11	**20**	18
snake	3	4	7	**20**	3	3	3	3	3
sokoban	6	13	**19**	10	8	11	12	10	9
spider	1	12	7	**20**	1	1	1	1	1
storage	14	18	**19**	**19**	17	15	**19**	18	**19**
termes	0	10	14	15	3	4	2	7	3
tetris	3	**20**	**20**	**20**	3	3	3	3	3
thoughtful	5	5	8	14	5	12	5	5	5
tidybot	2	19	16	**20**	2	3	3	4	2
tpp	6	13	22	**29**	6	15	10	14	15
transport	0	5	0	16	0	0	0	0	0
trucks	6	9	**15**	9	6	6	7	7	7
visitall	0	**20**	0	**20**	0	0	0	0	0
woodworking	1	1	2	4	1	1	1	1	1
zenotravel	3	**20**	**20**	**20**	3	9	13	14	13
SUM	359	775	933	965	346	603	558	681	632
$>h_{IC}$					4	6	4	6	7
$>h_{MT}$					2	5	3	4	3
$>h_{Min}$					0	2	1	0	0

Coverage	**arb**	**inv**	**rnd**	**GZD**	**BD**	**ZCA**	**VDM**	**ZCP**	**AM**	**GZD+BD**	**Scorpion**
Airport (50)	28	27	23	**29**	27	28	24	24	27	28	**29**
Blocks (35)	**28**	27	**28**	**28**	**28**	**28**	**28**	**28**	**28**	**28**	**28**
DataNetwork (20)	12	12	12	13	12	12	12	12	12	13	**14**
Depot (22)	7	7	7	7	7	7	7	7	7	10	**13**
DriverLog (20)	13	14	13	14	13	13	13	13	13	13	**15**
Elevators (30)	22	22	20	22	22	22	22	22	23	23	**24**
Freecell (80)	15	15	13	24	16	12	15	15	21	33	**64**
Grid (5)	2	2	1	2	2	2	2	2	2	2	**3**
Hiking (20)	10	10	9	10	9	9	9	9	10	10	**14**
Logistics (63)	27	27	25	27	25	25	25	25	25	27	**34**
Mprime (35)	23	25	22	23	23	22	22	22	25	24	**31**
Mystery (19)	16	17	15	17	17	17	17	17	17	17	**19**
Nomystery (20)	16	16	14	17	15	15	18	15	18	18	**20**
Openstacks (80)	31	31	31	31	31	30	31	31	31	31	**34**
OrgSynth-split (20)	**15**	**15**	14	**15**	14	10	**15**	**15**	**15**	**15**	10
Parcprinter (30)	19	22	19	22	19	19	22	22	20	20	**30**
Parking (40)	9	9	6	9	10	10	12	12	8	**13**	**13**
Pegsol (36)	**35**	34	33	**35**	34	34	**35**	34	34	**35**	**35**
Pipes-notank (50)	18	18	17	18	17	17	18	18	17	18	**25**
Pipes-tank (50)	12	12	10	12	11	9	12	12	12	12	**18**
PNetAlignment (20)	**9**	**9**	**7**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**0**
Rovers (40)	9	**11**	9	9	9	9	9	9	9	9	9
Satellite (36)	8	12	7	8	14	13	**15**	**15**	10	14	8
Scanalyzer (30)	16	16	16	16	15	14	16	16	16	16	**18**
Snake (20)	6	6	4	6	4	4	4	6	6	7	**13**
Sokoban (30)	**30**	**29**	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**
Spider (20)	11	11	9	12	11	11	11	11	10	12	**15**
Termes (20)	7	6	6	7	6	6	7	7	6	7	**13**
Tidybot (40)	**23**	22	20	**23**	20	15	22	22	22	**23**	22
VisitAll (40)	16	15	17	15	**36**	**36**	**36**	16	14	**36**	30
Woodworking (30)	19	22	19	19	20	22	20	20	22	20	**25**
Zenotravel (20)	**13**	**13**	12	**13**	12	12	**13**	**13**	**13**	12	**13**
Others (601)	331	331	331	331	331	331	331	331	331	331	**346**
Sum (1672)	856	865	821	873	869	850	884	883	858	914	**1020**

Empirical Evaluation – Examples from HSDIP'20

	Coverage			Time Score		
	offline	on-sample	on-suite	offline	on-sample	on-suite
agricola (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
airport (50)	**34**	**24**	**34**	2.5	19.5	**26.1**
barman (34)	4	0	4	0.3	0.0	**2.0**
blocks (35)	**28**	**20**	**28**	2.2	19.5	**29.1**
childsnack (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
data-network (20)	**14**	**11**	**14**	1.1	7.7	**10.2**
depot (22)	**13**	**6**	**13**	1.0	3.3	**9.6**
driverlog (20)	**15**	**7**	**15**	1.1	4.1	**10.7**
elevators (50)	**44**	**12**	**44**	3.4	2.6	**25.9**
floortile (40)	6	0	6	0.3	0.0	**0.8**
freecell (80)	**68**	**30**	**68**	**4.8**	**10.9**	**35.8**
ged (20)	**19**	**7**	**19**	1.5	4.7	**12.1**
grid (5)	3	1	3	0.2	1.0	**2.3**
gripper (20)	**8**	**6**	**8**	0.6	**4.8**	**7.0**
hiking (20)	**14**	**8**	**15**	1.0	**5.1**	**10.7**
logistics (63)	**39**	**19**	**39**	2.8	**11.8**	**25.2**
miconic (150)	**144**	**133**	**144**	**144**	**11.2**	**80.6**
movie (30)	**30**	**30**	**30**	**30**	**2.4**	**27.2**
mprime (35)	**29**	**24**	**29**	2.3	**19.1**	**26.5**
mystery (30)	**19**	**15**	**19**	1.5	**12.6**	**17.8**
mystery-prime (30)	**20**	**12**	**20**	1.5	**8.0**	**15.3**
openstacks (100)	**53**	**21**	**53**	**53**	3.8	**12.0**
organic (20)	7	7	7	0.5	6.0	**6.1**
organic-split (20)	**10**	**6**	**10**	0.7	1.9	**4.2**
parcprinter (30)	**38**	**34**	**38**	2.9	**28.2**	**34.1**
parking (40)	**13**	**1**	**13**	0.9	0.1	**5.4**
pathways (30)	5	4	5	0.3	3.1	**4.5**
pegsol (36)	**48**	**42**	**48**	**48**	3.7	**22.8**
petri-net (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
pipes-nt (50)	**25**	**14**	**25**	1.8	**10.4**	**19.0**
pipes-t (50)	**18**	**8**	**18**	1.3	5.1	**12.1**
pnt-small (50)	**50**	**49**	**50**	**50**	**3.9**	**48.2**
rovers (40)	**8**	**7**	**8**	0.6	6.6	**8.1**
satellite (36)	7	6	7	0.5	5.5	**7.2**
scanalyzer (50)	**35**	**7**	**33**	**35**	2.7	**5.7**
snake (20)	**12**	**6**	**12**	0.9	2.5	**7.6**
sokoban (50)	**50**	**33**	**50**	**50**	3.8	**19.6**
spider (20)	**15**	**7**	**15**	**15**	1.1	**2.9**
storage (30)	**16**	**14**	**16**	**16**	1.2	**12.5**
termes (20)	**12**	**0**	**12**	**12**	0.8	**0.0**
tetris (17)	**11**	**3**	**10**	**11**	0.8	1.3
tidybot (40)	**18**	**16**	**18**	**18**	1.8	**5.8**
tpp (30)	**8**	**7**	**8**	**8**	0.6	**8.0**
transport (70)	**34**	**20**	**36**	**36**	**2.6**	**10.4**
trucks (30)	**13**	**9**	**13**	**13**	0.9	**5.3**
visitall (40)	**30**	**33**	**30**	**30**	**2.3**	**27.8**
woodworking (50)	**49**	**58**	**49**	**49**	**3.8**	**24.3**
zenotravel (20)	**13**	**7**	**13**	**13**	1.0	**4.9**
Sum (1827)	1156	766	1155	**1159**	**86.8**	**539.8**

	blind	lama-first	ff	mercury	saarplan	SINGLE (Max)	CS	CS	CS	CS
agricola	0	0	10	16	1	13	12	11	9	
airport	23	35	36	33	1	21	13	14	17	
barman	0	0	12	20	0	0	0	0	0	
blocks	18	35	35	35	21	33	27	35	35	
childsnack	0	0	1	0	0	0	0	0	0	
data-network	0	0	5	2	0	0	0	0	0	
depot	4	14	18	18	6	5	5	15	13	
driverlog	7	19	18	18	9	10	15	15	14	
elevators	0	5	20	7	0	0	0	0	0	
floortile	0	0	1	1	0	0	0	0	0	
freecell	20	46	79	80	18	80	23	61	57	
ged	1	0	20	20	0	20	0	0	0	
grid	1	3	4	5	0	3	0	3	4	
gripper	8	20	20	20	8	20	10	20	20	
hiking	2	20	20	20	2	7	5	3	5	
logistics	2	7	29	15	2	2	3	6	5	
maintenance	0	14	11	14	0	0	0	0	0	
miconic	35	150	150	150	71	150	146	150	150	
movie	30	30	30	30	30	30	30	30	30	
mprime	20	21	32	22	5	18	13	18	17	
mystery	15	15	17	14	9	6	12	10	9	
openstacks	0	0	2	20	0	1	8	2	5	
organic-synthesis	3	3	3	3	3	3	3	3	2	
parcprinter	0	12	20	18	0	0	0	0	0	
parking	0	0	7	0	0	0	0	0	0	
pathways	4	5	10	4	3	4	4	4	4	
pegsol	17	20	20	20	18	20	20	20	20	
pipesworld	12	22	23	27	11	13	15	20	27	
psr	45	50	50	50	50	50	50	50	50	
rovers	6	21	26	25	6	14	16	16	16	
satellite	6	15	22	25	6	7	7	7	7	
scanalyzer	4	20	18	20	5	18	11	20	18	
snake	2	3	4	4	2	3	3	3	3	
sokoban	6	13	19	10	8	11	12	10	9	
spider	1	12	15	15	1	1	1	1	1	
storage	14	18	19	19	17	15	19	18	19	
termes	0	10	14	15	3	4	2	7	3	
tetris	20	20	20	20	5	3	4	5	5	
thoughtful	4	5	8	14	5	12	5	5	5	
tidybot	2	19	16	20	2	5	5	9	2	
tpp	6	13	22	29	6	15	10	14	15	
transport	0	5	0	16	0	9	0	0	0	
trucks	6	9	15	9	6	6	7	7	7	
visitall	0	20	0	20	0	0	9	0	0	
woodworking	1	1	2	4	1	1	1	1	1	
zenotravel	8	20	20	20	7	9	13	12	11	
SUM	359	775	933	965	346	603	558	631	632	
$>t_{\text{SF}}$					4	6	4	6	7	
$>t_{\text{SF}}$					2	5	2	5	6	
$>t_{\text{Min}}$					0	2	1	0	0	

Domain	#	Blind Search $\frac{\text{cov}}{\text{sc}}$	$\frac{\text{cov}}{\text{off}}$	Dominance Pruning $\frac{\text{cov}}{\text{sc}}$	$\frac{\text{cov}}{\text{off}}$	Duplicate Checking $\frac{\text{cov}}{\text{sc}}$	$\frac{\text{cov}}{\text{off}}$	A* with $h^{\text{LM-cut}}$ Dominance Pruning $\frac{\text{cov}}{\text{sc}}$	$\frac{\text{cov}}{\text{off}}$	Duplicate Checking $\frac{\text{cov}}{\text{sc}}$	$\frac{\text{cov}}{\text{off}}$
DataNet	20	**9**	**9**	5	5	9	9	5	5	5	5
Depots	22	3	3	4	4	4	4	7	7	7	7
Driverlog	20	**11**	**11**	**11**	**11**	**11**	**11**	9	9	10	10
Elevators	30	6	6	9	12	**16**	**16**	0	0	10	10
Floortile	40	**2**	**2**	**2**	**2**	**2**	**2**	0	0	0	0
Freecell	42	0	0	0	0	0	0	**2**	**2**	**2**	**2**
GED	20	13	13	**15**	**15**	**15**	**15**	**15**	**15**	**15**	**15**
Grid	5	1	1	1	1	1	1	**2**	**2**	**2**	**2**
Logistics	63	24	25	25	26	22	22	24	24	34	34
Miconic	145	**46**	**47**	**47**	**47**	45	45	42	42	135	135
NoMystery	20	**20**	**20**	**20**	**20**	**19**	**20**	**16**	**16**	**16**	**20**
OpenStk14	20	1	1	1	1	2	2	2	2	2	2
PSR	48	**48**	**48**	**46**	**48**	**48**	**48**	42	42	46	46
Rovers	40	7	7	7	7	6	6	7	7	8	8
Satellite	36	5	5	5	5	5	5	5	5	7	7
Tidybot14	20	1	3	3	3	3	3	3	3	**6**	**6**
Transport	28	10	11	13	14	**15**	**15**	0	0	13	13
Woodwork	26	7	7	7	7	7	7	16	16	16	16
Zenotravel	20	8	9	8	9	9	9	6	6	6	7
Others	239	67	67	67	67	67	67	67	67	87	87
Σ	894	291	296	296	309	310	**313**	242	242	285	288

Coverage	arb	inv	rnd	GZD	BD	ZCA	VDM	ZCP	AM	GZD+BD	Scorpion
Airport (50)	28	27	23	**29**	27	28	24	24	27	28	**29**
Blocks (35)	**28**	27	**28**	**28**	**28**	**28**	**28**	**28**	**28**	**28**	**28**
DataNetwork (20)	12	12	12	13	12	12	12	12	12	13	**14**
Depot (22)	7	7	7	7	7	7	7	7	7	7	**13**
DriverLog (20)	13	14	13	14	13	13	13	13	13	13	**15**
Elevators (30)	22	22	20	22	22	22	22	22	22	22	**24**
Freecell (80)	15	15	13	24	16	15	15	15	15	33	**64**
Grid (5)	2	2	1	2	2	2	2	2	2	2	**3**
Hiking (20)	10	10	9	10	10	9	10	10	10	10	**14**
Logistics (63)	27	27	25	27	25	25	25	25	25	27	**34**
Mprime (35)	23	25	22	23	23	22	23	23	25	24	**31**
Mystery (19)	16	17	15	17	17	17	17	17	17	17	**19**
Nomystery (20)	16	16	14	17	17	16	18	18	18	16	**20**
Openstacks (80)	31	31	31	31	31	30	31	31	31	31	**34**
OrgSynth-split (20)	**15**	**15**	14	**15**	14	10	**15**	**15**	**15**	15	10
Parcprinter (30)	19	22	19	22	19	19	22	22	22	22	**30**
Parking (40)	9	9	6	9	10	10	12	12	8	**13**	**13**
Pegsol (36)	**35**	34	33	**35**	34	34	34	34	34	34	**35**
Pipes-notank (50)	18	18	17	18	17	17	18	17	18	18	**25**
Pipes-tank (50)	12	12	10	12	11	12	12	12	12	12	**18**
PNetAlignment (20)	**9**	**9**	**7**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	0
Rovers (40)	9	**11**	9	9	9	9	9	9	9	9	9
Satellite (36)	8	12	7	8	14	13	**15**	**15**	10	14	8
Scanalyzer (30)	16	16	16	16	15	14	16	16	16	16	**18**
Snake (20)	6	6	4	6	4	4	6	6	6	7	**13**
Sokoban (30)	**30**	**29**	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**
Spider (20)	11	11	9	12	11	11	11	11	10	12	**15**
Termes (20)	7	6	6	7	6	6	7	7	6	7	**13**
Tidybot (40)	23	22	20	23	20	15	22	22	22	23	**32**
VisitAll (40)	16	15	17	15	**36**	**36**	**36**	16	14	**36**	30
Woodworking (30)	19	22	19	19	20	20	20	22	22	20	**30**
Zenotravel (20)	13	**13**	12	13	12	12	13	**13**	13	12	13
Others (601)	331	331	331	331	331	331	331	331	331	331	**346**
Sum (1672)	856	865	821	873	869	850	884	883	858	914	**1020**

Empirical Evaluation – Examples from HSDIP'20

	Coverage			Time Score		
	offline	on-node	on-sushi	offline	on-node	on-sushi
agricola (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
airport (50)	**34**	**24**	**34**	2.5	18.5	**26.1**
barman (34)	4	0	4	0.3	0.0	**2.0**
blocks (35)	**28**	**20**	**28**	2.2	19.5	**29.1**
childsnack (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
data-network (20)	**14**	**11**	**14**	1.1	7.2	**12.1**
depot (22)	**13**	**6**	**13**	1.0	3.3	**9.6**
diving (20)	**15**	**7**	**15**	1.1	4.1	**10.7**
elevators (50)	**44**	**12**	**44**	3.4	2.6	25.9
floortile (40)	6	0	6	0.3	0.0	**4.8**
freecell (80)	**68**	**30**	**68**	**68**	4.8	10.9
ged (20)	**19**	**7**	**19**	1.5	4.7	12.1
grid (5)	3	1	3	0.2	0.1	**2.3**
gripper (20)	**8**	**6**	**8**	**8**	0.6	**7.0**
hiking (20)	**14**	**8**	**15**	1.1	4.4	**10.5**
logistics (63)	**39**	**19**	**39**	2.8	11.8	25.2
miconic (150)	**144**	**133**	**144**	**144**	11.2	85.6
movie (30)	**30**	**30**	**30**	2.4	**42.7**	**42.4**
mprime (35)	**29**	**24**	**29**	2.3	19.1	26.5
mystery (30)	**19**	**15**	**19**	1.5	12.6	**17.8**
nomystery (20)	**20**	**12**	**20**	1.5	8.0	**15.8**
openstacks (100)	**53**	**21**	**53**	**53**	3.8	**12.0**
organic (20)	7	7	7	0.5	6.0	**6.1**
organic-split (20)	**10**	**6**	**10**	0.7	1.9	**4.2**
parcprinter (50)	**38**	**34**	**38**	**38**	2.9	28.2
parking (40)	**13**	**1**	**13**	0.9	0.1	**5.4**
pathways (30)	5	4	5	0.3	0.3	**5.4**
pegsol (50)	**48**	**42**	**48**	**48**	3.7	**22.8**
petri-net (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
pipes-nt (50)	**25**	**14**	**25**	**25**	1.8	10.4
pipes-t (50)	**18**	**8**	**18**	**18**	1.3	5.1
psr-small (50)	**50**	**49**	**50**	**50**	3.9	**48.2**
rovers (40)	**8**	**7**	**8**	**8**	0.6	6.6
satellite (36)	7	6	7	**7**	0.5	5.5
scanalyzer (50)	**35**	**7**	**33**	**35**	2.7	5.7
snake (18)	12	6	12	**12**	0.9	2.5
sokoban (50)	**50**	**33**	**50**	**50**	**50**	**39.6**
spider (20)	**15**	**7**	**15**	**15**	1.1	2.9
storage (30)	**16**	**14**	**16**	**16**	1.2	**12.5**
termes (20)	**12**	**0**	**12**	**12**	0.8	0.0
tetris (17)	**11**	**3**	**10**	**11**	0.8	1.3
tidybot (40)	**18**	**16**	**18**	**18**	1.8	5.8
tpp (30)	**8**	**7**	**8**	**8**	0.6	8.0
transport (70)	**34**	**20**	**36**	**36**	2.6	10.4
trucks (30)	**13**	**9**	**13**	**13**	0.9	5.3
visitall (40)	**30**	**33**	**30**	**30**	2.3	**27.8**
woodworking (50)	**49**	**58**	**49**	**49**	3.8	24.3
zenotravel (20)	**13**	**7**	**13**	**13**	1.0	**4.9**
Sum (1027)	1156	766	1155	**1159**	**86.6**	**539.9**

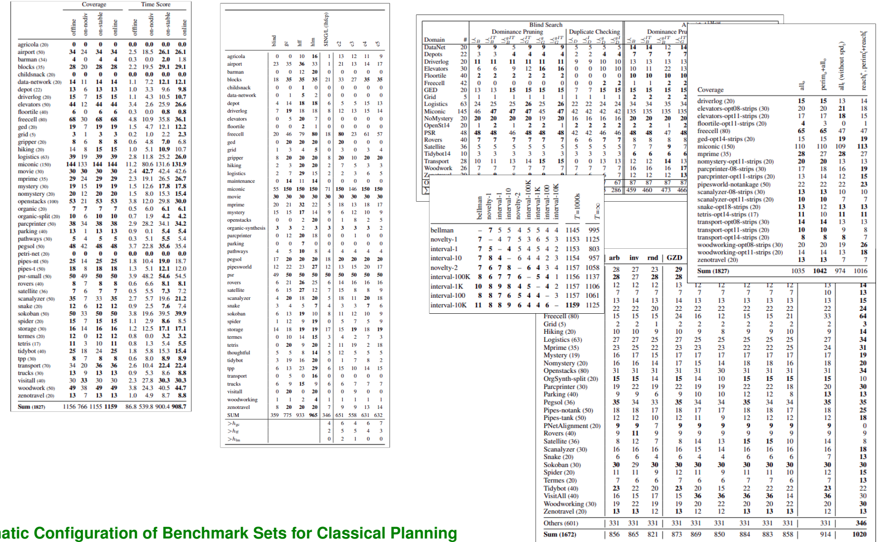
	Blind Search			A*			Dominance Pruning			Duplicate Checking			Dominance Pru		
Domain	#	$\leq_{st}$	$\leq_{lex}$	$\leq_{sto}$	$\leq_{lex}$	$\leq_{sto}^{pdt}$	$\leq_{lex}^{pdt}$	$\leq_{sto}^{sdt}$	$\leq_{lex}^{sdt}$	$\leq_{st}$	$\leq_{lex}$	$\leq_{sto}^{pdt}$	$\leq_{lex}^{pdt}$	$\leq_{sto}^{sdt}$	$\leq_{lex}^{sdt}$
DataNet	20	**9**	**9**	5	**9**	**9**	5	5	5	**14**	**14**	12	**14**
Depots	22	3	3	4	4	4	4	4	**7**	**7**	**7**	**7**	**7**
Driverlog	20	**11**	**11**	**11**	**11**	**11**	9	9	10	10	13	13	13
Elevators	30	6	6	9	12	**16**	9	9	10	10	10	11	22
Floortile	40	**2**	**2**	**2**	**2**	**2**	0	0	0	0	**10**	**10**	**10**
Freecell	42	0	0	0	0	0	0	0	2	2	1	1	2
GED	20	13	13	**15**	**15**	**15**	**15**	**15**	**15**	**15**	**15**	**15**	**15**
Grid	5	1	1	1	1	1	1	2	2	2	2	2	2
Logistics	63	24	25	25	26	25	22	22	24	24	34	35	34
Miconic	145	**46**	**47**	**47**	**47**	**47**	41	42	42	135	135	135	135
NoMystery	20	**20**	**20**	**20**	**20**	19	16	16	16	**20**	**20**	**20**	**20**
OpenStk14	20	2	2	2	2	2	2	2	2	2	2	2	2
PSR	48	**48**	**48**	46	**48**	**48**	48	42	46	46	**48**	**48**	47	**48**
Rovers	40	5	5	7	7	6	6	7	8	8	8	8	8	8
Satellite	36	5	5	5	5	5	5	5	5	7	7	9	7
Tidybot14	10	3	3	3	3	3	3	3	3	**6**	**6**	**6**	**6**
Transport	28	10	11	13	14	**15**	**15**	0	0	13	13	12	14	13
Woodwork	26	7	7	7	7	7	7	7	7	16	16	16	17	16
Zenotravel	20	8	9	8	9	9	9	6	6	7	12	12	12	12
Others	239	67	67	67	67	67	63	67	67	67	87	87	87	87
Σ	894	291	296	296	309	310	**313**	242	242	285	380	459	460	473	466

	total	ff	att	blin	SINGL (blin)	$\leq_1^1$	$\leq_1^2$	$\leq_2^1$	$\leq_2^2$
agricola	0	0	10	16	1	13	12	11	9
airport	23	35	**36**	33	1	21	13	14	17
barman	0	0	12	**20**	0	0	0	0	0
blocks	18	35	**35**	35	21	33	27	**35**	**35**
childsnack	0	0	1	0	0	0	0	0	0
data-network	0	1	5	2	0	0	0	0	0
depot	4	14	**18**	**18**	6	5	5	15	13
driverlog	7	**19**	18	18	9	15	15	14	14
elevators	0	5	**20**	7	0	0	0	0	0
floortile	0	0	2	0	0	0	0	0	0
freecell	20	46	79	**80**	18	**80**	23	61	57
ged	0	**20**	**20**	**20**	0	**20**	0	0	0
grid	1	3	4	5	0	3	3	3	4
gripper	8	**20**	**20**	**20**	8	**20**	**20**	**20**	**20**
hiking	2	**20**	**20**	**20**	2	7	5	3	5
logistics	2	7	**29**	15	2	2	3	6	5
maintenance	0	**14**	**14**	**14**	0	0	0	0	0
miconic	35	**150**	**150**	**150**	71	**150**	146	**150**	**150**
movie	30	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**
mprime	20	21	32	22	5	18	13	18	17
mystery	15	15	17	14	9	6	12	10	9
openstacks	0	0	2	**20**	0	1	0	2	0
organic-synthesis	3	3	3	3	3	3	3	3	2
parcprinter	12	12	**20**	18	1	0	0	0	0
parking	0	0	7	0	0	0	0	0	0
pathways	4	5	9	9	1	4	4	4	4
pegsol	17	**20**	**20**	**20**	18	**20**	**20**	**20**	**20**
pipesworld	12	22	23	**27**	12	13	15	20	**27**
psr	45	50	49	50	50	50	50	50	50
rovers	6	21	**26**	25	6	14	16	16	16
satellite	4	15	22	25	4	5	5	5	5
scanalyzer	4	**20**	18	**20**	5	18	11	**20**	18
snake	3	4	7	7	3	3	3	3	3
sokoban	6	13	19	10	8	11	12	10	9
spider	1	12	19	19	1	1	1	1	1
storage	14	18	19	19	17	15	19	18	19
termes	0	10	14	15	3	4	2	7	3
tetris	0	**20**	14	15	0	0	0	0	0
thoughtful	4	5	8	14	5	12	5	5	5
tidybot	3	19	16	**20**	3	9	7	9	9
tpp	6	13	22	**29**	6	15	10	14	15
transport	0	5	0	16	0	0	0	0	0
trucks	6	9	15	9	6	6	7	7	7
visitall	0	**20**	0	20	0	0	0	0	0
woodworking	1	1	2	4	1	1	1	1	1
zenotravel	3	**20**	**20**	**20**	3	9	13	14	13
SUM	359	775	933	965	346	651	558	631	632
$>_{tt}$					4	6	4	6	7
$>_{st}$					2	5	3	5	4
$>_{lex}$					0	2	1	0	0

Coverage	all	prntrs_wall	all_(without opt)	rmark_ prntrs(rmark)
driverlog (20)	**15**	**15**	13	14
elevators-opt08-strips (30)	20	**20**	**21**	**18**
elevators-opt11-strips (20)	17	17	**18**	15
floortile-opt11-strips (20)	4	3	3	3
freecell (80)	**65**	**65**	47	47
ged-opt14-strips (20)	15	15	**19**	**19**
miconic (150)	110	110	109	**113**
mprime (35)	**28**	**27**	**28**	**27**
nomystery-opt11-strips (20)	**20**	**20**	13	13
parcprinter-08-strips (30)	17	18	16	**19**
parcprinter-opt11-strips (20)	13	14	12	**15**
pipesworld-nontankage (50)	22	22	22	**23**
scanalyzer-08-strips (30)	**13**	**13**	**21**	10
scanalyzer-opt11-strips (20)	**10**	**10**	7	7
snake-opt18-strips (20)	13	**12**	13	12
tetris-opt14-strips (17)	**11**	**10**	**11**	**11**
transport-opt08-strips (30)	**14**	**14**	13	13
transport-opt11-strips (20)	**10**	**10**	9	8
transport-opt14-strips (20)	**9**	**8**	8	7
woodworking-opt08-strips (30)	20	20	19	**26**
woodworking-opt11-strips (20)	14	13	13	**18**
zenotravel (20)	**13**	**13**	7	7
Sum (1027)	1035	**1042**	974	1016

Coverage	arb	inv	rnd	GZD				
Airport (50)	28	27	23	**29**				
Blocks (35)	**28**	27	**28**	**28**				
DataNetwork (20)	12	12	12	13	12	12	12	12	13	**14**
Depot (22)	7	7	7	7	7	7	7	7	10	**13**
DriverLog (20)	13	14	13	14	13	13	13	13	13	**15**
Elevators (30)	22	22	20	22	22	22	22	22	22	**24**
Freecell (80)	15	15	13	24	16	15	15	15	21	**33**
Grid (5)	2	2	1	2	2	2	2	2	2	**3**
Hiking (20)	10	10	9	10	9	9	9	9	9	**14**
Logistics (63)	27	27	25	27	25	25	25	25	27	**34**
Mprime (35)	23	25	22	23	22	22	22	22	24	**31**
Mystery (19)	16	17	15	17	17	17	17	17	17	**19**
Nomystery (20)	16	16	14	17	17	16	18	18	16	**20**
Openstacks (80)	31	31	31	31	31	30	31	31	31	**34**
OrgSynth-split (20)	**15**	**15**	14	14	10	**15**	**15**	**15**	15	10
Parcprinter (30)	19	19	22	19	19	22	22	22	18	**20**
Parking (40)	9	9	6	9	10	10	12	12	8	**13**
Pegsol (50)	**35**	34	33	**35**	34	34	34	34	34	**35**
Pipes-notank (50)	18	18	17	18	17	17	18	18	17	**18**
Pipes-tank (50)	12	12	10	12	12	12	12	12	12	**13**
PNetAlignment (20)	**9**	**9**	**7**	**9**	**9**	**9**	**9**	**9**	**9**	**9**
Rovers (40)	9	**11**	9	9	9	9	9	9	9	9
Satellite (36)	8	12	7	8	14	13	**15**	**15**	10	14
Scanalyzer (30)	16	16	16	16	15	14	16	16	16	**18**
Snake (20)	6	6	4	6	4	4	4	4	6	**8**
Sokoban (30)	**30**	**29**	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**
Spider (20)	11	11	9	12	11	11	11	11	10	**12**
Termes (20)	7	6	6	7	6	6	7	7	6	**7**
Tidybot (40)	**23**	**22**	**20**	**23**	**22**	**22**	**22**	**22**	**23**	**22**
VisitAll (40)	16	15	17	15	**36**	**36**	**36**	14	**36**	30
Woodworking (30)	19	22	19	19	20	20	20	22	20	**23**
Zenotravel (20)	**13**	**13**	12	13	12	12	**13**	**13**	12	**13**
Others (601)	331	331	331	331	331	331	331	331	331	**346**
Sum (1672)	856	865	821	873	869	850	884	883	858	**914**	**1020**

Empirical Evaluation – Examples from HSDIP'20



Empirical Evaluation – Examples from HSDIP'20

Coverage

	Time Score						
	offline	on-sat01	on-sat08	online	offline	on-sat01	on-sat08	online
agricola (26)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
airport (50)	**34**	**24**	**34**	**34**	2.5	19.5	**29.1**	**26.1**
barman (34)	4	0	4	4	0.3	0.0	**2.0**	1.8
blocks (35)	**28**	**20**	**28**	**28**	2.2	19.5	**29.1**	**26.1**
childsnack (20)	**0**	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**	**0.0**
data-network (20)	**14**	**11**	**14**	**14**	1.1	7.2	**12.1**	**12.1**
depot (22)	**13**	**6**	**13**	**13**	1.0	3.3	9.6	**9.8**
driverlog (20)	**15**	**7**	**15**	**15**	1.1	4.3	**12.5**	**10.7**
elevators-sat08 (30)	**44**	**12**	**44**	**44**	3.4	2.6	25.9	**26.6**
floortile (40)	**6**	0	**6**	**6**	0.3	0.0	**4.8**	**4.8**
freecell (80)	**68**	**30**	**68**	**68**	4.8	10.9	**35.8**	**36.1**
ged (20)	**19**	**7**	**19**	**19**	1.5	4.7	**12.1**	**12.2**
grid (5)	**3**	1	**3**	**3**	0.2	0.1	**2.2**	**2.3**
gripper (20)	**8**	**6**	**8**	**8**	0.6	4.8	**7.0**	**6.8**
hiking (20)	**14**	**8**	**15**	**15**	1.0	5.1	**10.5**	**10.7**
logistics (63)	**39**	**19**	**39**	**39**	2.8	11.8	25.2	**26.0**
miconic (150)	**144**	**133**	**144**	**144**	11.2	80.6	131.9	**134.9**
movie (30)	**30**	**30**	**30**	**30**	2.4	**42.7**	42.4	42.6
mprime (35)	**29**	**24**	**29**	**29**	2.3	19.1	26.5	**26.7**
mystery (30)	**19**	**15**	**19**	**19**	1.5	12.6	**17.8**	**17.8**
nomystery-sat11 (20)	**20**	**12**	**20**	**20**	1.5	8.0	15.3	**18.2**
openstacks (100)	**53**	**21**	**53**	**53**	3.8	12.0	29.8	**30.0**
organic (20)	7	7	7	7	0.5	**4.0**	**4.1**	**4.1**
organic-split (20)	**10**	**6**	**10**	**10**	0.7	1.9	**4.2**	**4.2**
parcprinter (50)	**38**	**34**	**38**	**38**	2.9	28.2	34.1	**34.2**
parking (40)	**13**	**1**	**13**	**13**	0.9	0.1	**5.4**	**5.4**
pathways (30)	5	4	5	5	0.3	3.2	**5.0**	**5.4**
pegsol (48)	**48**	**42**	**48**	**48**	3.7	**22.8**	**35.6**	**35.4**
petri-net (20)	**0**	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**	**0.0**
pipes-nt (50)	**25**	**14**	**25**	**25**	1.8	10.0	18.7	**19.2**
pipes-t (50)	**15**	/			ds		ss	
psr-small (50)	**50**	**49**			+sib	+sub	+sib	+sub
rovers (40)	**8**	/						
satellite (36)	7	/	airport (50)		**25**	**25**	**24**	**24**
scanalyzer (50)	**35**	/	data-network (20)		13	13	**14**	**14**
snake (20)	**12**	/	freecell (80)		43	60	**61**	**61**
sokoban (50)	**50**	**X**	ged (20)		15	15	**19**	**19**
spider (20)	**15**	/	hiking (20)		11	11	12	12
storage (30)	**16**	/	logistics98 (35)		12	12	12	12
termes (20)	**11**	/	miconic (150)		**144**	**143**	**144**	**144**
tetris (17)	**11**	/	mprime (35)		30	30	30	30
tidybot (40)	**19**	/	mystery (30)		18	18	**19**	**19**
tpp (30)	**8**	/	openstacks-08 (30)		20	20	22	22
transport (58)	34	**X**	openstacks-11 (20)		15	15	17	17
trucks (30)	**13**	/	openstacks-14 (20)		3	3	3	3
visitall (40)	30	**X**	parcprinter-08 (30)		**30**	**30**	**28**	**28**
woodworking (30)	**29**	**X**	parcprinter-11 (20)		**20**	**20**	18	18
zenotravel (20)	13	/	parking-11 (20)		4	4	7	7
			parking-14 (20)		4	4	6	6
Sum (827)	1156	782	pipesworld-not. (50)		21	21	**24**	**24**
			pipesworld-t. (50)		14	13	**16**	**16**
			rovers (40)		**11**	**11**	10	10
			scanalyzer-08 (30)		15	15	**18**	**18**
			scanalyzer-11 (20)		12	12	**15**	**15**
			snake (20)		4	4	10	10
			spider (20)		12	12	14	**15**
			termes (20)		10	10	12	12
			tetris (17)		8	8	10	**10**
			tidybot-11 (20)		12	13	**14**	**14**
			tidybot-14 (20)		4	4	8	**9**
			transport (20)		10	10	13	13
			sum (827)		536	537	596	596
			other domains (-sub)		**535**	**535**	**535**	**535**
			total (1027)		1071	1072	1131	1131

Blind Search

Domain	#	$\frac{\#\sigma}{\#\pi}$	$\frac{\#\sigma}{\#\pi^*}$	$\frac{\#\sigma}{\#\pi^{\mathrm{GED}}}$	$\frac{\#\sigma}{\#\pi^{\mathrm{FD}}}$	$\frac{\#\sigma}{\#\pi^{\mathrm{BJ}}}$	$\frac{\#\sigma}{\#\pi^{\mathrm{SINGL}}}$	$\frac{\#\sigma}{\#\pi^{\mathrm{blind}}}$
DataNet	20	**9**	**9**	**5**	**9**	**9**	**5**	**5**
Depots	22	3	3	4	5	2	**4**	**4**
Driverlog	20	**11**	**11**	**11**	**11**	**11**	**9**	**9**
Elevators	30	6	6	**9**	**16**	**16**	**9**	**9**
Floortile	40	**2**	**2**	**2**	**2**	**2**	**0**	**0**
Freecell	42	0	0	0	0	0	**2**	**2**
GED	20	13	13	**15**	**15**	**15**	**15**	**15**
Grid	5	1	1	1	1	**2**	**2**	**2**
Logistics	63	24	25	25	**26**	**25**	**22**	**24**
Miconic	145	**46**	**47**	**47**	**47**	**47**	**41**	**42**
NoMystery	20	**20**	**20**	**20**	**20**	**19**	**16**	**16**
OpenStk14	20	**2**	**2**	**2**	**2**	**2**	**2**	**2**
PSR	48	**48**	**48**	**46**	**48**	**48**	**42**	**46**
Rovers	40	5	5	7	7	7	**8**	**8**
Satellite	36	5	5	5	5	5	5	5
Tidybot14	10	3	3	3	3	3	3	3
Transport	28	10	11	13	14	**15**	**15**	**0**
Woodwork	26	7	7	7	7	7	7	7
Σ		219	222	232	249	246	214	203
GED		219	222	232	249	246	214	203

Bidirectional Search

Domain	$\frac{\#\sigma}{\#\pi}$	$\frac{\#\sigma}{\#\pi^*}$	$\frac{\#\sigma}{\#\pi^{\mathrm{GED}}}$	$\frac{\#\sigma}{\#\pi^{\mathrm{FD}}}$	$\frac{\#\sigma}{\#\pi^{\mathrm{BJ}}}$	$\frac{\#\sigma}{\#\pi^{\mathrm{SINGL}}}$	$\frac{\#\sigma}{\#\pi^{\mathrm{blind}}}$
DataNet	**9**	**9**	**5**	**9**	**9**	**5**	**5**
Depots	22	3	3	4	5	2	**4**
Driverlog	**20**	**11**	**11**	**11**	**11**	**9**	**9**
Elevators	30	6	6	**9**	**16**	**16**	**9**
Floortile	40	**2**	**2**	**2**	**2**	**2**	**0**
Freecell	42	0	0	0	0	0	**2**
GED	20	13	13	**15**	**15**	**15**	**15**
Grid	5	1	1	1	1	**2**	**2**
Logistics	63	24	25	25	**26**	**25**	**22**
Miconic	145	**46**	**47**	**47**	**47**	**47**	**41**
NoMystery	20	**20**	**20**	**20**	**19**	**16**	**16**
OpenStk14	20	**2**	**2**	**2**	**2**	**2**	**2**
PSR	48	**48**	**48**	**46**	**48**	**48**	**42**
Rovers	40	5	5	7	7	7	**8**
Satellite	36	5	5	5	5	5	5
Tidybot14	10	3	3	3	3	3	3
Transport	28	10	11	13	14	**15**	**15**
Woodwork	26	7	7	7	7	7	7
Σ		219	222	232	249	246	214
GED		219	222	232	249	246	214

Domain	#	bellman	novelty-1	interval-1	novelty-2	interval-10	interval-100	interval-1K	interval-10K	Σ t=1000s	Σ t=∞
bellman	–	7	5	5	4	5	5	4	4	1145	995
novelty-1	7	–	4	7	5	3	6	5	3	1153	1125
interval-1	7	5	–	4	4	4	5	4	4	1137	1125
interval-10	7	8	4	–	6	4	4	2	3	1154	957
novelty-2	7	6	7	8	–	6	4	3	4	1157	1058
interval-100K	7	6	7	8	5	–	5	4	5	1157	1137
interval-1K	10	8	9	8	4	5	–	4	2	1157	1106
interval-10K	10	8	8	7	6	5	4	–	3	1157	1061
interval-100	11	8	8	8	9	6	4	4	–	1159	1125

Domain	bfsf	ff	iiff	ihsp	SINGL (bfs)	C1	C2	C3	C4	C5
agricola	0	0	10	16	1	13	12	11	9	
airport	23	35	**36**	33	1	21	13	14	17	
barman	0	0	12	**20**	0	0	0	0	0	
blocks	18	35	**35**	35	21	33	27	**35**	**35**	
childsnack	0	0	1	0	0	0	0	0	0	
data-network	0	1	5	2	0	0	0	0	0	
depot	4	14	**18**	**18**	6	5	5	15	13	
driverlog	7	**19**	**19**	**19**	10	15	15	14		
elevators	0	5	**20**	7	0	0	0	0	0	
floortile	0	0	0	0	0	0	0	0	0	
freecell	20	46	**79**	**80**	**18**	**80**	23	61	57	
ged	1	0	**20**	**20**	0	**20**	0	0	0	
grid	1	4	4	3	1	3	3	3	4	
gripper	4	**20**	**20**	**20**	8	**20**	**20**	**20**		
hiking	2	**20**	**20**	**20**	2	7	5	3	5	
logistics	2	7	**29**	15	2	2	3	6	5	
maintenance	0	**14**	11	14	0	0	0	0	0	
miconic	35	**150**	**150**	**150**	71	**150**	146	**150**	**150**	
movie	30	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**	
mprime	20	21	32	**35**	5	10	13	16	17	
mystery	15	15	17	14	9	6	12	10	9	
openstacks	0	0	2	**20**	0	1	0	2	5	
organic-synthesis	**3**	**3**	2	3	**3**	**3**	**3**	3	2	
parcprinter	4	12	12	12	4	4	4	4	4	
parking	0	0	7	0	0	0	0	0	0	
pathways	5	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	
pegsol	17	**20**	**20**	**20**	**18**	**20**	**20**	**20**	**20**	
	12	22	23	**27**	12	13	15	20	17	
	45	**50**	**50**	**50**	**50**	**50**	**50**	**50**	**50**	
	6	21	**26**	25	6	14	16	18	16	
	4	7	7	7	4	7	7	7	7	
	4	20	18	**20**	5	18	11	**20**	18	
	0	0	0	0	0	0	0	0	0	
	6	13	**19**	10	8	11	12	10	9	
	11	11	12	12	**13**	12	12	12	12	
	14	18	**19**	**19**	17	15	**19**	18	**19**	
	0	10	14	15	3	4	2	7	3	
	0	**20**	19	15	0	0	0	0	0	
	3	5	8	14	5	12	5	5	5	
	2	19	**20**	**20**	2	3	3	4	2	
	6	13	22	**29**	6	15	10	14	15	
	0	0	0	0	0	0	0	0	0	
	6	9	15	9	6	6	7	7	7	
	0	0	**20**	0	0	0	0	0	0	
	1	1	2	4	1	1	1	1	1	
	1	**20**	**20**	**20**	9	13	13	13	13	
sum	375	775	933	965	346	631	558	631	632	
					4	6	4	6	7	
					2	3	2	3	3	
					0	2	1	0	0	

Coverage	all	pettis_wall1	all_(without opt1)
driverlog (20)	**15**	**15**	13
elevators-opt08-strips (30)	20	**20**	**21**
elevators-opt11-strips (20)	17	17	**18**
floortile-opt11-strips (20)	4	3	**5**
freecell (80)	**65**	**65**	47
ged-opt14-strips (20)	15	15	**19**
miconic (150)	110	110	**109**
mprime (35)	**28**	**27**	**28**
nomystery-opt11-strips (20)	**20**	**20**	13
parcprinter-08-strips (30)	17	18	**19**
parcprinter-opt11-strips (20)	13	14	**15**
pipesworld-notankage (50)	22	22	**22**
pipesworld-tankage (50)	**13**	**13**	**10**
scanalyzer-opt11-strips (20)	**10**	**7**	7
snake-opt18-strips (20)	13	**12**	8
tetris-opt14-strips (17)	**11**	**10**	**11**
transport-opt08-strips (30)	**14**	**14**	**13**
transport-opt11-strips (20)	**10**	**9**	8
transport-opt14-strips (20)	**9**	**8**	7
woodworking-opt08-strips (30)	20	20	**19**
woodworking-opt11-strips (20)	14	14	**18**
zenotravel (20)	**13**	**13**	7
Sum (827)	1035	**1042**	974
			1016

Domain	#	$\frac{\#\sigma}{\#\pi}$	$\frac{\#\sigma}{\#\pi^*}$	$\frac{\#\sigma}{\#\pi^{\mathrm{GED}}}$	$\frac{\#\sigma}{\#\pi^{\mathrm{FD}}}$	$\frac{\#\sigma}{\#\pi^{\mathrm{BJ}}}$	$\frac{\#\sigma}{\#\pi^{\mathrm{SINGL}}}$	$\frac{\#\sigma}{\#\pi^{\mathrm{blind}}}$
Freecell (80)		15	15	13	14	14	13	13
Grid (5)		2	2	1	2	2	2	2
Hiking (20)		10	10	10	10	10	10	10
Logistics (63)		27	27	25	27	25	25	25
Mprime (35)		23	23	22	22	22	22	22
Mystery (19)		16	17	15	17	17	17	17
Nomystery (20)		16	16	14	17	15	18	18
Openstacks (80)		31	31	31	31	31	31	31
OrgSynth-split (20)		**15**	**15**	14	**15**	**15**	**15**	**15**
Parcprinter (50)		19	22	19	22	22	22	18
Parking (40)		9	9	6	9	10	12	12
Pegsol (36)		**35**	**35**	34	**35**	34	**35**	34
Pipes-notank (50)		18	18	17	18	17	17	18
Pipes-tank (50)		12	12	10	12	12	12	12
PNetAlignment (20)		**9**	**9**	**7**	**9**	**9**	**9**	**9**
Rovers (40)		9	11	9	9	9	9	9
Satellite (36)		8	12	7	8	14	13	**15**
Scanalyzer (30)		16	16	16	16	15	14	16
Snake (20)		6	6	4	4	4	6	6
Sokoban (30)		**30**	**29**	**30**	**30**	**30**	**30**	**30**
Spider (20)		11	11	9	12	11	11	11
Termes (20)		7	6	6	7	6	7	7
Tidybot (40)		**23**	**22**	**20**	**20**	19	**22**	**23**
VisitAll (40)		16	15	17	**35**	**36**	**36**	14
Woodworking (30)		12	12	12	12	12	12	12
Zenotravel (20)		**13**	**13**	12	**13**	**13**	**13**	12
Others (601)		331	331	331	331	331	331	331
Sum (1672)		856	865	821	873	869	850	884

	C1	C2	C3	C4	C5
	14	14	12	14	
	2	2	2	2	
	10	10	10	10	
	25	25	25	25	
	23	22	22	22	
	17	17	17	17	
	18	18	18	18	
	31	31	31	31	
	15	**15**	**15**	**15**	
	22	19	22	22	
	10	10	12	12	
	35	34	34	34	
	17	17	17	18	
	12	12	12	12	
	9	**9**	**9**	**9**	
	9	9	9	9	
	13	13	13	13	
	16	16	16	16	
	6	6	6	6	
	30	**30**	**30**	**30**	
	11	11	11	11	
	7	7	7	7	
	23	**23**	**22**	**22**	
	14	**36**	**36**	14	
	12	12	12	12	
	13	**13**	**13**	**13**	
	331	331	331	331	**346**
	869	850	884	883	858

	C1	C2	C3	C4	C5
	13	14			
	2	2			
	10	10			
	25	25			
	22	22			
	17	17			
	18	18			
	31	31			
	15	**15**			
	22	22			
	13	**13**			
	34	34			
	18	18			
	12	12			
	9	**9**			
	9	9			
	15	**15**			
	16	16			
	6	6			
	30	**30**			
	11	11			
	7	7			
	23	**23**			
	14	**36**			
	12	12			
	13	**13**			
	331	331			**346**
	869	850	884	883	858

	C1	C2	C3	C4	C5
	13	14			
	2	2			
	10	10			
	25	25			
	22	22			
	17	17			
	18	18			
	31	31			
	15	**15**			
	22	22			
	13	**13**			
	34	34			
	18	18			
	12	12			
	9	**9**			
	9	9			
	15	**15**			
	16	16			
	6	6			
	30	**30**			
	11	11			
	7	7			
	23	**23**			
	14	**36**			
	12	12			
	13	**13**			
	331	331			**346**
	869	850	884	883	858
	914	**1020**			

Empirical Evaluation – Examples from HSDIP'20

	Coverage			Time Score		
	offline	on-sample	on-unkn	offline	on-sample	on-unkn
agricola (26)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
airport (50)	**34**	**24**	**34**	2.5	19.5	**26.1**
barman (34)	4	0	4	0.3	0.0	**2.0**
blocks (35)	**28**	**20**	**28**	2.2	19.5	**29.1**
childsnack (20)	**0**	**0**	**0**	**0.0**	**0.0**	**0.0**
data-network (38)	**14**	**11**	**14**	1.1	7.5	**12.1**
depot (22)	**13**	**6**	**13**	1.0	3.3	**9.6**
driverlog (20)	**15**	**7**	**15**	1.1	4.3	**10.7**
elevators-sat08 (30)	**44**	**12**	**44**	3.4	2.6	**25.9**
floortile (40)	4	0	4	0.3	0.0	**0.8**
freecell (80)	**68**	**30**	**68**	**4.8**	**10.9**	**35.8**
ged (20)	**19**	**7**	**19**	1.5	4.7	**12.1**
grid (5)	3	1	**3**	0.2	0.7	**2.3**
gripper (20)	**8**	**6**	**8**	0.6	**4.8**	**7.0**
hiking (20)	**14**	**8**	**15**	1.0	**7.1**	**10.5**
logistics (63)	**39**	**19**	**39**	2.8	**11.8**	**25.2**
miconic (150)	**144**	**133**	**144**	**144**	11.2	**85.6**
movie (30)	**30**	**30**	**30**	**30**	**24.2**	**42.7**
mprime (35)	**29**	**24**	**29**	**29**	**21.9**	**26.5**
mystery (30)	**19**	**15**	**19**	**19**	1.5	**12.6**
nomystery-sat11 (20)	**20**	**12**	**20**	**20**	1.5	**15.3**
openstacks (100)	**53**	**21**	**53**	**53**	**3.8**	**12.0**
organic (20)	7	7	7	7	0.5	**4.0**
organic-split (20)	**10**	**6**	**10**	**10**	0.7	**1.9**
parcprinter (50)	**38**	**34**	**38**	**38**	2.9	**28.2**
parking (40)	**13**	**1**	**13**	**13**	0.9	**0.1**
pathways (30)	5	4	5	5	0.3	**5.5**
pegsol (48)	**48**	**42**	**48**	**48**	**3.7**	**22.8**
petri-net (20)	**0**	**0**	**0**	**0**	**0.0**	**0.0**
pipes-nt (50)	**25**	**14**	**25**	**25**	**1.8**	**10.0**
pipes-t (50)	**15**	/				
pnl-small (36)	**50**	**49**				
rovers (40)	**8**	/				
satellite (36)	7	/				
scanalyzer (50)	**35**	/				
snake (20)	**12**	/				
sokoban (50)	**50**	**X**				
spider (20)	**15**	/				
storage (30)	**16**	/				
termes (20)	**1**	/				
tetris (17)	**11**	/				
tidybot (40)	**18**	/				
tpp (30)	**8**	/				
transport (70)	**34**	**X**				
trucks (30)	**13**	/				
visitall (40)	**30**	**X**				
woodworck. (30)	**49**	**X**				
zenotravel (20)	**13**	/				
Sum (1827)	1156	**76**				

	Blind	IA	AIT	Bliss	SINGUL (blks)
agricola	0	0	10	16	1
airport	23	35	**36**	33	1
barman	0	0	12	**20**	0
blocks	18	32	**35**	35	21
childsnack	0	0	1	0	0
data-network	1	1	5	2	0
depot	4	14	**18**	**18**	6
driverlog	2	**19**	**19**	**19**	0
elevators	0	5	**20**	7	0
floortile	0	0	0	0	0
freecell	20	46	**79**	**80**	**18**
ged	1	0	**20**	**20**	0
grid	1	2	4	4	0
gripper	4	**20**	**20**	**20**	**8**
hiking	2	3	**20**	**20**	2
logistics	2	7	**29**	15	2
maintenance	0	**14**	11	10	0
miconic	35	**150**	**150**	**150**	**71**
movie	30	**30**	**30**	**30**	**30**
mprime	20	21	32	**22**	5
mystery	15	15	**17**	16	9
openstacks	0	1	2	**20**	0
organic-synthesis	3	3	3	3	3
parcprinter	6	12	**20**	12	0
parking	0	0	7	0	0
pathways	0	0	0	0	0
pegsol	17	**28**	**26**	/	/
	12	22	23	/	/
	6	21	**26**	/	/
	4	6	6	/	/
	4	7	**20**	**18**	/
	6	13	**19**	/	/
	3	12	12	/	/
	14	**18**	**19**	/	/
	0	10	14	/	/
	1	**20**	**20**	/	/
	3	5	8	/	/
	6	9	**15**	/	/
	6	13	23	/	/
	1	3	4	/	/
	1	**20**	**20**	/	/
359	775	933	**1035**	1035	1035

Blind Search				Dominance Pruning				Duplicate Checking				Dominance Pru			
Domain	#	$\subseteq_\#$	$\subseteq_\$$	$\subseteq_\#^{\text{eff}}$	$\subseteq_\$^{\text{eff}}$	$\subseteq_\#^{\text{eff}}$	$\subseteq_\$^{\text{eff}}$	$\subseteq_\#$	$\subseteq_\$$	$\subseteq_\#^{\text{eff}}$	$\subseteq_\$^{\text{eff}}$	$\subseteq_\#$	$\subseteq_\$$	$\subseteq_\#^{\text{eff}}$	$\subseteq_\$^{\text{eff}}$
DataNet	20	**9**	**9**	5	5	**9**	**9**	5	5	5	5	**14**	**14**	**12**	**14**
Depots	22	3	3	4	4	4	4	4	4	7	7	7	7	7	7
Driverlog	20	**11**	**11**	**11**	**11**	**11**	**11**	9	9	10	10	13	13	13	13
Elevators	30	6	6	9	9	**16**	**16**	0	0	10	10	10	11	22	23
Floortile	40	**2**	**2**	**2**	**2**	**2**	**2**	0	0	0	0	**10**	**10**	**10**	**10**
Freecell	42	0	0	0	0	0	0	0	0	2	2	1	1	2	2
GED	20	13	13	**15**	**15**	**15**	**15**	7	7	**15**	**15**	**15**	**15**	**15**	**15**
Grid	5	1	1	2	2	2	2	1	1	2	2	2	2	2	2
Logistics	63	24	25	25	26	25	**26**	22	22	24	24	34	34	35	34
Miconic	145	**47**	**47**	**47**	**47**	**47**	**47**	42	42	46	46	135	135	135	135
NoMystery	20	**20**	**20**	**20**	**20**	19	**20**	16	16	16	16	**20**	**20**	**20**	**20**
OpenStack	20	1	1	2	2	2	2	1	1	2	2	2	2	2	2
PSR	48	**48**	**48**	46	46	**48**	**48**	42	42	46	46	**48**	**48**	47	**48**
Rovers	40	7	7	7	7	7	7	6	6	7	7	8	8	8	8
Satellite	36	5	5	5	5	5	5	5	5	5	5	7	7	9	7
Tidybot14	10	3	3	3	3	3	3	3	3	3	3	**6**	**6**	**6**	**6**
Transport	28	10	11	13	14	**15**	**15**	0	0	13	13	12	12	14	13
Woodwork	26	7	7	7	7	7	7	7	7	7	7	16	16	16	17
∑		7	7	7	7	7	7	7	7	7	7	12	12	12	13
CR		47	47	47	47	47	47	47	47	47	47	47	47	47	47
		280	459	460	473	466	**502**	329	329	401	413	580	603	615	636

	bellman	novelty-1	novelty-2	interval-10	novelty-3	interval-100	interval-1K	interval-10K	$T_\#$=1000s	T_{SCC}
bellman	–	7	5	5	4	5	4	4	1145	995
novelty-1	7	–	4	7	5	3	6	5	1153	1125
interval-1	7	5	–	9	6	5	4	4	1153	1125
	4	–	6	4	4	2	3	1154	957	

Collection	Comp	PDBs	Scorp	LM-cut	SymBA*
bipartite	105	73	**182**	180	59
bd-bipartite	96	**78**	**137**	107	72
chain 0.5	558	321	**1004**	372	282
chain 0.5	349	327	**389**	359	288
chain 0.25	409	345	**444**	428	309
chain 0.75	320	304	**391**	333	261
complete	123	117	**153**	129	110
dag 0.1	113	94	**179**	166	77
dag 0.2	65	45	**108**	**121**	25
dag 0.25	66	50	**134**	108	15
dag 0.75	73	62	**113**	110	32
dichain	382	241	**433**	296	249
fork	350	321	**421**	395	308
inverted fork	**436**	393	356	386	337
layer 0.25	224	224	**384**	343	224
polytree 0.5	323	269	**373**	358	254
polytree 0.25	367	268	**405**	374	295
polytree 0.75	276	**408**	408	408	260
random 0.1	54	38	**167**	95	32
random 0.5	57	52	**117**	88	34
random 0.25	97	89	**153**	129	83
random 0.75	56	39	**81**	50	35
star 0.1	226	130	**272**	224	125
star 0.5	257	144	**386**	272	139
star 0.25	219	118	**290**	205	115
star 0.75	172	132	**263**	199	134
tree	364	277	**433**	345	245
	4,993	4,014	7,233	5,902	4416

Domain	#	ada	portfo_woSL	portfo_wSL	rmaS_portfo	rmaS_ada	ada	portfo_woSL	portfo_wSL	rmaS_portfo	rmaS_ada
Coverage											
driverlog (20)		**15**	**15**	13	14						
elevators-opt08-strips (30)		20	**20**	18	18						
elevators-opt11-strips (20)		17	17	**18**	15						
floortile-opt11-strips (20)		4	3	3	**5**						
freecell (80)		**65**	65	47	47						
ged-opt14-strips (20)		15	15	**19**	19						
miconic (150)		110	110	109	**113**						
mprime (35)		**28**	27	28	27						
nomystery-opt11-strips (20)		**20**	**20**	13	13						
parcprinter-08-strips (30)		17	18	16	**19**						
parcprinter-opt11-strips (20)		13	14	12	**15**						
pipesworld-notankage (50)		22	22	22	**23**						
scanalyzer-08-strips (30)		**13**	**13**	10	10						
scanalyzer-opt11-strips (20)		**10**	**10**	7	7						
snake-opt14-strips (20)		**13**	12	13	12						
termes-opt14-strips (17)		**11**	**10**	**11**	**11**						
transport-opt11-strips (20)		**14**	**14**	13	13						
transport-opt14-strips (20)		**10**	**10**	9	8						
visitall-opt14-strips (20)		**8**	**8**	5	7						
woodworking-opt08-strips (30)		20	20	19	**26**						
woodworking-opt11-strips (20)		14	14	13	**18**						
zenotravel (20)		**13**	**13**	7	7						
Sum (1827)		1035	**1042**	974	1016						

Domain	#	arb	inv	rnd	GZD
DataNet	20	28	27	23	**29**
Depots	22	**28**	27	**28**	**28**
Driverlog	20	12	12	12	13
Elevators	30	7	7	7	7
Floortile	40	2	2	2	2
Freecell	80	15	15	13	24
Grid	5	2	2	2	2
Hiking	20	10	10	10	10
Logistics	63	27	27	25	25
Mprime	35	23	23	22	22
Mystery	19	16	17	15	17
Nomystery	20	16	14	17	18
Openstacks	80	31	31	31	31
OrgSynth-split	20	**15**	**15**	**15**	**15**
Parcprinter	30	19	22	22	18
Parking	40	9	9	6	9
Pegsol	36	**35**	34	34	**35**
Pipes-notank	50	18	17	17	18
Pipes-tank	50	12	10	12	12
PNetAlignment	20	**9**	**9**	**9**	**9**
Rovers	40	8	8	9	8
Satellite	36	8	12	7	8
Scanalyzer	30	16	16	16	15
Snake	20	6	4	4	6
Sokoban	30	**30**	**29**	**30**	**30**
Spider	20	11	9	12	11
Termes	20	7	6	6	7
Tidybot	40	23	22	20	22
VisitAll	40	15	17	15	**36**
Woodworking	30	19	21	19	19
Zenotravel	20	**13**	**13**	12	12
Others (601)		331	331	331	331
Sum (1672)		856	865	821	873

	arb	inv	rnd	GZD
	13	13	13	**14**
	7	7	7	10
	13	13	13	**15**
	22	22	22	22
	23	23	22	23
	2	2	2	3
	10	10	10	10
	25	25	25	25
	22	22	22	22
	17	17	17	17
	18	18	18	16
	31	31	31	31
	15	**15**	**15**	**15**
	22	22	18	22
	10	10	12	8
	34	34	34	**35**
	17	18	17	18
	12	12	12	12
	9	**9**	**9**	**9**
	9	9	9	9
	7	7	7	7
	14	13	15	14
	16	16	16	16
	4	4	4	6
	30	**30**	**30**	**30**
	11	11	11	11
	7	7	7	7
	22	22	22	22
	15	15	**36**	**36**
	19	19	19	19
	12	12	13	**13**
	331	331	331	331
	873	869	850	884

	arb	inv	rnd	GZD
	331	331	331	**346**
	883	858	883	914

	arb	inv	rnd	GZD
	13	13	13	**14**
	10	10	10	13
	13	13	13	**15**
	22	22	22	22
	23	23	22	23
	3	3	3	3
	10	10	10	10
	25	25	25	25
	22	22	22	22
	17	17	17	17
	18	18	18	16
	31	31	31	31
	15	**15**	**15**	**15**
	22	22	18	22
	10	10	12	8
	34	34	34	**35**
	17	18	17	18
	12	12	12	12
	9	**9**	**9**	**9**
	9	9	9	9
	7	7	7	7
	14	13	15	14
	16	16	16	16
	4	4	4	6
	30	**30**	**30**	**30**
	11	11	11	11
	7	7	7	7
	22	22	22	22
	15	**36**	**36**	**36**
	19	19	19	19
	12	12	13	**13**
	331	331	331	331
	873	869	850	884

	arb	inv	rnd	GZD
	331	331	331	**346**
	883	858	883	914

Empirical Evaluation – The ICAPS/IPC Way

The ICAPS/IPC Way

- Measure coverage
- Time limit 30 minutes
- Memory limit 2-8 GB
- Use the benchmarks from the International Planning Competition

Empirical Evaluation – The ICAPS/IPC Way

The ICAPS/IPC Way

- Measure coverage
- Time limit 30 minutes
- Memory limit 2-8 GB
- Use the benchmarks from the International Planning Competition

Having a standard evaluation setting is generally beneficial:

- Reproducibility
- Interpretability
- Avoids hand picking results

Empirical Evaluation – The ICAPS/IPC Way

The ICAPS/IPC Way

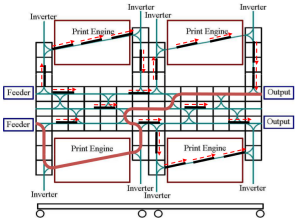
- Measure coverage
- Time limit 30 minutes
- Memory limit 2-8 GB
- Use the benchmarks from the International Planning Competition

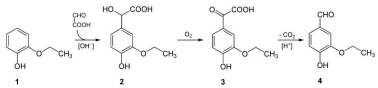
Having a standard evaluation setting is generally beneficial:

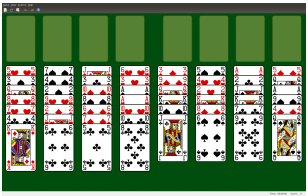
- Reproducibility
- Interpretability
- Avoids hand picking results

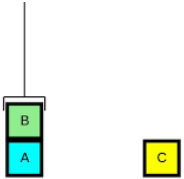
Outline

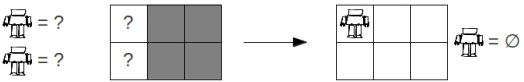
The diversity in the IPC Benchmark Set

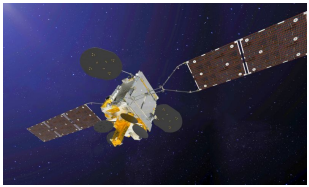












So, What's Wrong with the IPC Benchmark Set?

	IPC		
	L	D	O
Nomystery (20)	11	**20**	12
Rovers (40)	**40**	**40**	**40**
Woodworking (50)	**50**	**50**	**50**
Total	101	**110**	102

Table: Coverage of LAMA (L), Decstar (D) and OLCFF (O)

So, What's Wrong with the IPC Benchmark Set?

	IPC		
	L	D	O
Nomystery (20)	11	**20**	12
Rovers (40)	**40**	**40**	**40**
Woodworking (50)	**50**	**50**	**50**
Total	101	**110**	102

Table: Coverage of LAMA (L), Decstar (D) and OLCFF (O)

- Different number of instances per domain
- Instance scaling: too easy, too hard, and not smooth

So, What's Wrong with the IPC Benchmark Set?

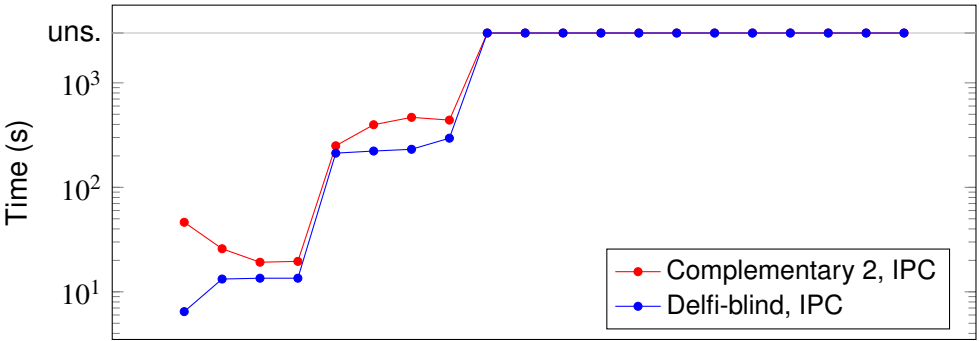
	IPC			New'14		
	L	D	O	L	D	O
Nomystery (20)	11	**20**	12	25	**30**	24
Rovers (40)	**40**	**40**	**40**	**22**	18	21
Woodworking (50)	**50**	**50**	**50**	18	27	**30**
Total	101	**110**	102	65	**75**	**75**

Table: Coverage of LAMA (L), Decstar (D) and OLCFF (O)

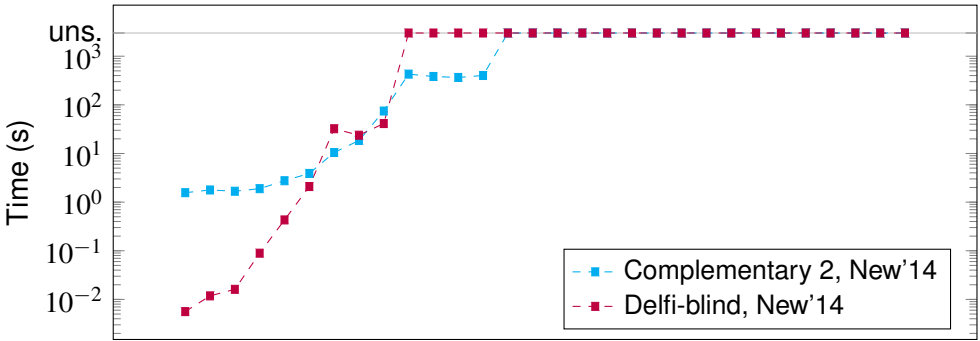
- Different number of instances per domain
- Instance scaling: too easy, too hard, and not smooth

→Experiments on some domains of the IPC benchmark set may not observe any difference between planners even if it exists!

Non-Smooth Scaling



Smooth Scaling



Contribution

An automatic tool to select instances from a given domain (more informative than the IPC set to compare current and future planners)

Contribution

An automatic tool to select instances from a given domain (more informative than the IPC set to compare current and future planners)

1. Smooth scaling from easy to hard instances:

Contribution

An automatic tool to select instances from a given domain (more informative than the IPC set to compare current and future planners)

1. Smooth scaling from easy to hard instances:
   - Easy: solvable by any planner that anyone would compare against (baseline)
   - Hard: out of reach of current existing planners within a reasonable time limit

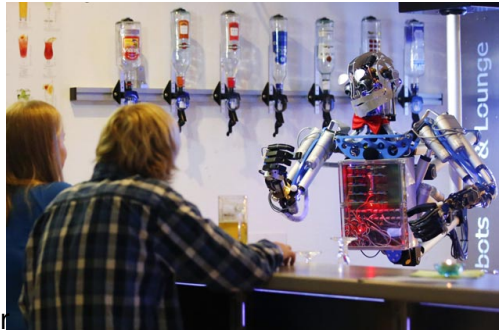
Contribution

An automatic tool to select instances from a given domain (more informative than the IPC set to compare current and future planners)

1. Smooth scaling from easy to hard instances:
   - Easy: solvable by any planner that anyone would compare against (baseline)
   - Hard: out of reach of current existing planners within a reasonable time limit
2. Minimize bias towards/against planners used

Outline

Example Domain: Barman



Instance Generator

```
./barman-generator.py <num_cocktails> <num_ingredients>
                      <num_shots> [<random_seed>]
  num_cocktails (min 1)
  num_ingredients (min 2)
  num_shots (min num_cocktails+1)
  random_seed (min 1, set to 2018)
```

Instance Generation Problem

Input:

- domain
- instance generator
- a baseline planner
- a set of state-of-the-art planners

Output: set of instances with a good scaling

Instance Generation Problem

Input:

- domain
- instance generator
- a baseline planner
- a set of state-of-the-art planners

Output: set of instances with a good scaling

Generate instances $\rightarrow$ Compute/Estimate runtimes $\rightarrow$ Select instances

Instance Generation Problem

Input:

- domain
- instance generator
- a baseline planner
- a set of state-of-the-art planners

Output: set of instances with a good scaling

Generate instances $\rightarrow$ Compute/Estimate runtimes $\rightarrow$ Select instances

How to avoid bias wrt. the set of considered planners?

Instance Generation Problem

Input:

- domain
- instance generator
- a baseline planner
- a set of state-of-the-art planners

Output: ~~set of instances~~ with a good scaling

~~Generate instances → Compute/Estimate runtimes → Select instances~~

How to avoid bias wrt. the set of considered planners?

Output: set of linear scaling of parameters for the generator that produce a good scaling in runtime

Sequences of instances

User specifies characteristics of the generator parameters:

Linear Attributes

cocktails	shots	ingredients
$b \in [1,6]$ $m \in [1,5]$	$b \in [1,5]$ $m \in [0,5]$ + cocktails	$v \in \{3,4,5\}$

Enumerated attributes

Sequences of instances

User specifies characteristics of the generator parameters:

Linear Attributes:

- Numeric value
- Increase size of the task
- User specifies ranges for the base value (b) and slope (m)

Enumerated attributes

cocktails	shots	ingredients
$b \in [1,6]$ $m \in [1,5]$	$b \in [1,5]$ $m \in [0,5]$ + cocktails	$v \in \{3,4,5\}$

Sequences of instances

User specifies characteristics of the generator parameters:

Linear Attributes:

- Numeric value
- Increase size of the task
- User specifies ranges for the base value (b) and slope (m)

Enumerated attributes:

- Finite set of values
- Fixed in the sequence

cocktails	shots	ingredients
$b \in [1,6]$ $m \in [1,5]$	$b \in [1,5]$ $m \in [0,5]$ + cocktails	$v \in \{3,4,5\}$

Sequences of instances

User specifies characteristics of the generator parameters:

Linear Attributes:

- Numeric value
- Increase size of the task
- User specifies ranges for the base value (b) and slope (m)

Enumerated attributes:

- Finite set of values
- Fixed in the sequence

cocktails	shots	ingredients
$b \in [1,6]$ $m \in [1,5]$	$b \in [1,5]$ $m \in [0,5]$ + cocktails	$v \in \{3,4,5\}$

Our system may select sequences like:

$(b = 5,$ $m = 1.34)$	$(b = 1, m = 0,$ +cocktails)	$(v = 3)$
5	6	3
6	7	3
7	8	3
9	10	3
10	11	3
11	12	3
13	14	3

Optimization Process

1. Generate candidate sequences that scale smoothly
2. Choose selected (sub-)sequences to include easy to hard instances

Sequence Optimization

We use SMAC to optimize the value of b, m and v for each parameter

- Measure instance difficulty as the best runtime by any planner (time limit 180 seconds)
- Penalty based on smoothness of scaling difficulty (ideally by a factor of 1.5 to 2)

Runtimes:	10.36, 15.41, 18.9, 28.02, 29.27, 68.01
Ratios:	1.48, 1.22, 1.48, 1.04, 2.32
Penalties:	0.02, 0.54, 0.02, 0.91, 0.13
Total penalty	1.62

Sequence Selection

MIP encoding to select sequences satisfying hard constraints:

- There are 30 instances
- (Easy) Baseline solves at least one instance in less than 30 seconds
- (Hard) Sub-sequences go from easy ($\leq 180s$) to hard ($>$ 2000s)
- (Diverse) Don't repeat the same parameters more than twice

and soft constraints:

- (Easy) Baseline solves 2 to 6 instances under 30 seconds
- (Easy) State-art planners solve 8 to 15 instances under 180 seconds
- (Hard) All sequences end in a very hard instance
- (Diverse) Don't repeat the same parameters more than once
- (Smooth) Minimize penalty of selected sequences

Outline

Experiments

Compare our new benchmark sets against the IPC

- 26 domains
- Satisficing and Optimal track
- 2 new benchmark sets that differ on the “training set”:
  - New’14: using planners up to 2014
  - New’20: using all available planners
- Evaluation based on planners from IPC’18

Evaluation Criteria

How to evaluate the quality of a benchmark set?

Evaluation Criteria

How to evaluate the quality of a benchmark set?

- Coverage range: generally better if all planners solve some instance and no planner solves all instances
- Comparisons: number of pairs (X, Y) of planners, such that $coverage(X) \neq coverage(Y)$

Evaluation Criteria

How to evaluate the quality of a benchmark set?

- Coverage range: generally better if all planners solve some instance and no planner solves all instances
- Comparisons: number of pairs (X, Y) of planners, such that $coverage(X) \neq coverage(Y)$

Goodhart's law: "When a measure becomes a target, it ceases to be a good measure." – Marilyn Strathern

$\rightarrow$Comparisons is a useful metric to compare benchmarks but not a metric to optimize for (would introduce bias towards the set of planners)

Results

		coverage range			comparisons		
Optimal	#IPC	IPC	'14	'20	IPC	'14	'20
barman	34	4–11	9–13	9–12	12	**21**	19
blocksworld	35	18–30	5–12	5–12	18	**24**	**24**
childsnack	20	0–6	9–20	6–21	12	18	**22**
data-network	20	6–14	5–12	5–16	**27**	25	**27**
depot	22	5–14	9–25	8–16	**26**	**26**	24
driverlog	20	7–15	6–30	5–18	22	**26**	25
elevators	50	28–44	7–14	10–18	**26**	**26**	23
floortile	40	16–34	9–18	8–17	21	21	**22**
grid	5	1–3	6–26	4–21	19	**28**	27
gripper	20	8–20	11–30	11–30	**7**	**7**	**7**
hiking	20	12–18	7–9	5–16	23	15	**25**
logistics	63	13–34	5–17	5–14	**27**	**27**	25
miconic	150	56–142	4–28	3–30	25	27	**28**
nomystery	20	8–20	3–27	5–21	18	**28**	27
openstacks	130	42–71	4–11	3–7	**24**	18	7
parking	40	0–15	11–18	12–21	**28**	24	23
rovers	40	6–13	4–26	6–19	**25**	22	7
satellite	36	7–14	8–30	4–27	22	25	**26**
scanalyzer	50	21–33	6–16	7–15	**27**	24	24
snake	20	7–14	5–20	7–19	22	**24**	21
storage	30	15–18	9–25	2–19	21	**27**	26
tpp	30	7–20	7–30	2–7	**24**	**24**	21
transport	70	24–35	5–30	8–19	21	18	**22**
visitall	40	12–30	6–21	5–20	**27**	**27**	**27**
woodworking	50	38–50	16–25	10–14	22	**26**	24
zenotravel	20	8–13	6–30	3–13	23	26	**28**

		coverage range			comparisons		
Satisficing	#IPC	IPC	'14	'20	IPC	'14	'20
barman	40	39–40	7–25	9–30	7	24	**27**
blocksworld	35	35–35	7–24	4–22	0	27	**28**
childsnack	20	1–20	14–30	2–19	27	25	**28**
data-network	20	9–19	10–30	13–30	24	**27**	25
depot	22	21–22	12–20	11–26	7	**27**	22
driverlog	20	20–20	29–30	9–19	0	12	**24**
elevators	50	49–50	30–30	30–30	**7**	0	0
floortile	40	4–40	1–12	1–11	17	**25**	24
grid	5	5–5	4–20	9–21	0	**26**	24
gripper	20	20–20	26–30	26–30	0	**7**	**7**
hiking	20	10–20	2–22	3–26	24	**28**	27
logistics	63	51–63	5–30	5–26	17	**27**	26
miconic	150	150–150	30–30	30–30	**0**	**0**	**0**
nomystery	20	12–20	19–30	2–30	23	18	**26**
openstacks	160	99–160	12–21	14–23	21	**27**	25
parking	40	36–40	14–20	13–16	7	**24**	21
rovers	40	38–40	10–22	6–30	7	26	**27**
satellite	36	26–36	5–30	6–14	**23**	17	**23**
scanalyzer	50	48–50	9–16	13–14	12	**21**	12
snake	20	3–17	6–30	5–14	27	**28**	26
storage	30	21–30	6–26	7–17	26	**27**	26
tpp	30	29–30	10–26	6–21	15	**27**	**27**
transport	70	65–70	22–30	15–23	7	24	**26**
visitall	40	36–40	4–30	4–29	7	24	**26**
woodworking	50	28–50	6–30	5–30	13	**27**	**27**
zenotravel	20	20–20	6–29	5–17	0	23	**25**

Highlight: SAT track

	comparisons		
Satisficing	IPC	'14	'20
gripper	0	7	7
miconic	0	0	0
elevators	7	0	0
blocksworld	0	27	28
driverlog	0	12	24
grid	0	26	24
zenotravel	0	23	25
barman	7	24	27
depot	7	27	22
parking	7	24	21
rovers	7	26	27
transport	7	24	26
visitall	7	24	26

Outline

Conclusion

- New tool to automatically select instances
  - Our tool consistently generates well-scaled instance sets that are useful to evaluate current planners
- New benchmark set significantly better than the IPC benchmark set, specially in the SAT/AGL track

Conclusion

- New tool to automatically select instances
  - Our tool consistently generates well-scaled instance sets that are useful to evaluate current planners
- New benchmark set significantly better than the IPC benchmark set, specially in the SAT/AGL track
- We need your feedback!
  - Do you find the results of our tool useful?
  - Is there any reason to prefer the IPC set over our new one?
  - Are there any constraints that we should take into account (in general or for specific domains)?