ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/273640829

Analysing differences between algorithm configurations through ablation

Article *in* Journal of Heuristics · January 2015

DOI: 10.1007/s10732-014-9275-9

CITATIONS
29

READS
42

2 authors:

Chris Fawcett
University of British Columbia - Vancouver
14 PUBLICATIONS **143** CITATIONS
SEE PROFILE

Holger H Hoos
Leiden University
292 PUBLICATIONS **14,633** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Computer Go View project

Energy Supply Cooperative (ESC) View project

All content following this page was uploaded by Holger H Hoos on 20 March 2017.

The user has requested enhancement of the downloaded file.

Journal of Heuristics

Analysing differences between algorithm configurations through ablation

Electronic supplementary material – Online resource 1

Chris Fawcett · Holger H. Hoos

This document presents the full ablation path data for every experimental scenario contained in our paper, in all cases where we decided not to include these tables in the main text.

Table 1 contains the training and test set performance, as well as [10%, 90%] bootstrap confidence intervals on the test set, for each configuration on the ablation path between the SPEAR default and the optimized configuration found by PARAMILS. Table 2 contains the same data for ablation performed between the CPLEX default and the optimized configuration, and Tables 3, 4 and 5 contain the same data for ablation performed between the LPG default and the optimized configurations for the depots, satellite and zenotravel instance sets, respectively.

Next are tables corresponding to our experiments performing ablation analysis between two high quality configurations. Table 6 contains the modified parameters, test set performance and [10%, 90%] bootstrap confidence intervals for configurations on the ablation path obtained by performing ablation on SPEAR between the best and second-best configurations found by PARAMILS. Table 7 contains the same data, for ablation on CPLEX between the best and

C. Fawcett
University of British Columbia
Computer Science Department
201-2366 Main Mall
Vancouver, BC, V6T 1Z4, Canada
Tel.: +1 604 822-4599
Fax: +1 604 822-5485
E-mail: fawcettc@cs.ubc.ca

H.H. Hoos
University of British Columbia
Computer Science Department
201-2366 Main Mall
Vancouver, BC, V6T 1Z4, Canada
Tel.: +1 604 822-1964
Fax: +1 604 822-5485
E-mail: hoos@cs.ubc.ca

second-best configurations found by ParamILS. Tables 8, 9 and 10 contain the same data for LPG on the depots, satellite and zenotravel instance sets, respectively.

We then include tables for our experiments performing ablation analysis for LPG between high-quality configurations for different problem domains, evaluated on one of those domains. Each table contains the parameters flipped, test set performance and [10%, 90%] bootstrap confidence interval for each configuration along the ablation path between the source and target configuration. First, Table 11 contains the results for ablation performed using the depots optimised configuration as the source and the satellite optimised configuration as the target, evaluated on the depots instance set. Table 12 has the equivalent data for ablation between the satellite and depots optimised configurations, evaluated on the satellite instance set. Tables 13 and 14 contain data for ablation between the depots and zenotravel optimised configurations, evaluated on the depots and zenotravel instance sets respectively. Lastly Tables 15 and 16 contain the data for ablation performed between the satellite and zenotravel optimised configurations, respectively evaluated on the satellite and zenotravel instance sets.

Next are tables containing the results for ablation analysis performed for LPG between high-quality configurations for different problem domains, this time evaluated on a third problem domain. Table 17 contains data for ablation between the depots and satellite optimised configurations, evaluated on the zenotravel instance set. Table 18 contains the data for ablation between the depots and zenotravel optimized configurations, evaluated on the satellite instance set. The remaining Table 19 contains the data for ablation between the satellite and zenotravel optimized configurations, evaluated on the depots instance set.

Finally, we present results for ablation with our improved support for conditional parameters, using LPG as the target algorithm. In each case the tables contain the modified parameters, training and test set performance, and [10%, 90%] bootstrap confidence intervals for each configuration along the ablation path between the LPG default and the configuration obtained by ParamILS. Tables 20, 21 and 22 contain the data for ablation using the depots, satellite and zenotravel configurations as the target, respectively.

Table 1 Parameters chosen and resulting PAR10 performance on both the training and test sets for ablation analysis accelerated by racing (maximum of 200 rounds) for SPEAR on SWV training set. Ablation was performed using the SPEAR default configuration as the source, and the configuration produced by PARAMILS as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance		Test bootstrap CI
		Training	Test	[10%, 90%]
default	-	573.649	569.645	[482.236, 658.573]
1	sp-var-dec-heur	2.041	1.800	[1.432, 2.190]
2	sp-res-cutoff-cls	2.341	1.930	[1.517, 2.359]
3	sp-rand-var-dec-scaling	1.554	1.287	[1.041, 1.538]
4	sp-use-pure-literal-rule	1.350	1.276	[1.041, 1.522]
5	sp-rand-var-dec-freq	1.365	1.288	[1.050, 1.541]
6	sp-var-activity-inc	1.428	1.325	[1.060, 1.613]
7	sp-clause-decay	1.413	1.310	[1.050, 1.586]
8	sp-learned-size-factor	1.420	1.302	[1.048, 1.581]
9	sp-res-cutoff-lits	1.432	1.319	[1.051, 1.609]
10	sp-rand-phase-dec-freq	1.405	1.308	[1.058, 1.582]
11	sp-clause-activity-inc	1.412	1.302	[1.050, 1.570]
12	sp-orig-clause-sort-heur	1.418	1.341	[1.061, 1.650]
13	sp-learned-clauses-inc	1.442	1.311	[1.056, 1.593]
14	sp-max-res-runs	1.404	1.331	[1.086, 1.589]
15	sp-learned-clause-sort-heur	1.352	2.018	[1.299, 2.857]
16	sp-first-restart	1.273	1.214	[0.996, 1.454]
17	sp-restart-inc	1.476	1.452	[1.118, 1.818]
18	sp-res-order-heur	1.488	1.327	[1.085, 1.572]
19	sp-variable-decay	1.417	1.377	[1.100, 1.660]
20	sp-phase-dec-heur	12.038	1.602	[1.292, 1.922]
21	sp-max-res-lit-inc	1.359	1.321	[1.055, 1.617]
configured	-	1.359	1.321	[1.055, 1.617]

Table 2 Parameters chosen and resulting PAR10 performance on both the training and test sets for ablation analysis accelerated by racing (maximum of 200 rounds) for CPLEX on the CORLAT training set. Ablation was performed using the CPLEX default configuration as the source, and the configuration produced by PARAMILS as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance		Test bootstrap CI
		Training	Test	[10%, 90%]
default	-	556.531	471.722	[429.845, 515.609]
1	mip_cuts_covers	90.776	63.067	[47.885, 78.801]
2	mip_strategy_heuristicfreq	16.599	21.385	[14.206, 29.822]
3	simplex_dgradient	8.080	7.368	[6.563, 8.205]
4	simplex_tolerances_markowitz	10.345	12.763	[7.103, 18.718]
5	mip_limits_aggforcut	6.900	6.120	[5.475, 6.798]
6	mip_strategy_variableselect	8.974	8.012	[4.800, 11.432]
7	simplex_pgradient	5.799	8.037	[4.815, 11.477]
8	mip_strategy_fpheur	8.699	5.155	[4.647, 5.687]
9	lpmethod	8.700	5.151	[4.654, 5.679]
10	barrier_crossover	8.662	5.146	[4.654, 5.671]
11	preprocessing_symmetry	8.696	5.140	[4.641, 5.657]
12	sifting_algorithm	8.695	5.145	[4.635, 5.672]
13	barrier_limits_growth	8.678	5.137	[4.638, 5.660]
14	mip_limits_gomorycand	8.704	5.143	[4.641, 5.670]
15	mip_limits_cutsfactor	5.639	5.115	[4.635, 5.614]
16	mip_cuts_gubcovers	5.672	5.132	[4.650, 5.633]
17	mip_cuts_gomory	5.447	5.236	[4.727, 5.760]
18	preprocessing_repeatpresolve	5.608	5.433	[4.948, 5.938]
19	mip_strategy_presolvenode	5.766	5.288	[4.830, 5.769]
20	mip_cuts_mircut	5.511	5.411	[4.923, 5.922]
configured	-	5.511	5.411	[4.923, 5.922]

Table 3 Parameters chosen and resulting PAR10 performance on both the training and test sets for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the depots training set. Ablation was performed using the LPG default configuration as the source, and the configuration produced by PARAMILS as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance		Test bootstrap CI
		Training	Test	[10%, 90%]
default	-	43.245	38.097	[30.224, 46.138]
1	cri_intermediate_levels	1.610	1.520	[1.430, 1.614]
2	triomemory	1.482	1.370	[1.278, 1.464]
3	donot_try_suspected_actions	1.460	1.345	[1.255, 1.436]
4	walkplan	1.310	1.162	[1.093, 1.236]
5	weight_mutex_in_relaxed_plan	1.069	1.012	[0.948, 1.079]
6	noise	1.143	1.033	[0.973, 1.093]
7	static_noise	1.124	1.067	[1.000, 1.135]
8	ri_list	1.132	1.069	[1.000, 1.138]
9	numrestart	1.117	1.062	[0.996, 1.127]
10	maxnoise	1.114	1.062	[0.998, 1.129]
11	inc_re	1.086	1.008	[0.949, 1.069]
12	extended_effects_evaluation	1.167	1.096	[1.030, 1.163]
13	numtry	1.035	1.013	[0.949, 1.078]
14	no_cut	1.035	1.015	[0.950, 1.082]
15	vicinato	1.038	1.014	[0.950, 1.082]
16	hpar_cut_neighb	0.671	0.658	[0.622, 0.696]
configured	-	0.671	0.658	[0.622, 0.696]

Table 4 Parameters chosen and resulting PAR10 performance on both the training and test sets for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the satellite training set. Ablation was performed using the LPG default configuration as the source, and the configuration produced by PARAMILS as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance		Test bootstrap CI
		Training	Test	[10%, 90%]
default	-	17.962	17.940	[17.846, 18.036]
1	cri_intermediate_levels	7.316	7.577	[7.538, 7.616]
2	criprecond	6.967	7.141	[7.102, 7.180]
3	noise	7.029	6.997	[6.959, 7.033]
4	no_cut	6.925	6.803	[6.769, 6.837]
5	improve_reachability	6.943	6.837	[6.802, 6.871]
6	donot_try_suspected_actions	6.834	6.793	[6.758, 6.827]
7	numrestart	6.915	6.776	[6.742, 6.809]
8	dynoisecoefnum	6.859	6.787	[6.753, 6.820]
9	cri_update_iterations	6.856	6.795	[6.761, 6.829]
10	vicinato	6.887	6.788	[6.753, 6.823]
11	hpar_cut_neighb	5.357	5.317	[5.287, 5.345]
12	triomemory	5.357	5.325	[5.297, 5.353]
13	verifyinit	5.337	5.329	[5.300, 5.356]
14	noise_incr	5.356	5.351	[5.323, 5.379]
15	maxnoise	5.471	5.343	[5.314, 5.371]
16	evaluate_mutex_for_action_remotion	5.360	5.325	[5.297, 5.354]
17	walkplan	5.363	5.340	[5.312, 5.368]
18	evaluation_function	5.217	5.339	[5.311, 5.367]
19	dynoiseTabLen	5.239	5.341	[5.313, 5.370]
20	static_noise	5.314	5.328	[5.299, 5.356]
21	relaxed_examination	5.268	5.314	[5.287, 5.341]
22	reset_extended_unsupported_facts	5.413	5.331	[5.302, 5.359]
23	mutex_and_additive_effects	5.423	5.344	[5.316, 5.371]
24	numtry	5.460	5.339	[5.310, 5.368]
25	not_supported_preconds_evaluation	5.355	5.335	[5.307, 5.364]
26	inc_re	5.379	5.327	[5.299, 5.355]
27	zero_num_A	5.331	5.335	[5.306, 5.364]
28	not_extended_unsupported_facts	5.372	5.344	[5.317, 5.372]
29	fast_best_action_evaluation	5.371	5.341	[5.312, 5.369]
30	no_mutex_with_additive_effects	5.375	5.331	[5.303, 5.357]
31	lagrange	5.376	5.334	[5.306, 5.362]
32	h	5.662	5.783	[5.752, 5.814]
configured	-	5.662	5.783	[5.752, 5.814]

Table 5 Parameters chosen and resulting PAR10 performance on both the training and test sets for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the zenotravel training set. Ablation was performed using the LPG default configuration as the source, and the configuration produced by PARAMILS as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance		Test bootstrap CI
		Training	Test	[10%, 90%]
default	-	29.671	29.361	[28.946, 29.775]
1	fast_best_action_evaluation	8.064	8.258	[8.189, 8.327]
2	triomemory	5.789	5.912	[5.858, 5.968]
3	noise	4.266	4.369	[4.314, 4.424]
4	criprecond	3.701	3.713	[3.685, 3.742]
5	stop_remove_act	3.458	3.446	[3.419, 3.474]
6	verifyinit	3.418	3.435	[3.408, 3.463]
7	weight_mutex_in_relaxed_plan	3.393	3.421	[3.394, 3.449]
8	reset_extended_unsupported_facts	3.377	3.401	[3.372, 3.429]
9	inc_re	3.348	3.387	[3.360, 3.415]
10	twalkplan	3.347	3.392	[3.364, 3.420]
11	choose_min_numA_fact	3.357	3.391	[3.364, 3.419]
12	numrestart	3.375	3.393	[3.365, 3.421]
13	lagrange	3.376	3.391	[3.364, 3.419]
14	verifyAf	3.371	3.392	[3.364, 3.420]
15	no_cut	3.368	3.397	[3.368, 3.425]
16	zero_num_A	3.362	3.393	[3.365, 3.421]
17	noise_incr	3.333	3.388	[3.360, 3.415]
18	numtry	3.401	3.389	[3.361, 3.416]
19	bestfirst	3.372	3.389	[3.361, 3.417]
20	evaluation_function	3.343	3.388	[3.360, 3.416]
21	no_pruning	3.367	3.391	[3.363, 3.419]
22	h	3.385	3.392	[3.364, 3.419]
23	donot_try_suspected_actions	3.367	3.388	[3.360, 3.416]
24	cri_insertion_add_mutex	3.371	3.387	[3.359, 3.414]
25	vicinato	3.370	3.381	[3.354, 3.409]
26	hpar_cut_neighb	2.035	2.047	[2.021, 2.073]
27	nonuniform_random	2.018	2.056	[2.029, 2.082]
28	dynoiseTabLen	1.996	2.053	[2.027, 2.079]
29	no_insert_threated_act_in_neighb	2.005	2.052	[2.025, 2.079]
30	maxnoise	2.065	2.092	[2.066, 2.119]
configured	-	2.065	2.092	[2.066, 2.119]

Table 6 Parameters chosen and resulting test set performance for ablation analysis accelerated by racing (maximum of 200 rounds) for SPEAR on SWV test set, using the best PARAMILS configuration as the source and the second-best PARAMILS configuration as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance	Test bootstrap CI [10%, 90%]
best	-	1.321	[1.055, 1.617]
1	sp-res-cutoff-lits	1.899	[1.058, 2.824]
2	sp-learned-clauses-inc	11.104	[1.056, 21.274]
3	sp-max-res-lit-inc	1.734	[1.391, 2.096]
4	sp-phase-dec-heur	1.316	[1.084, 1.564]
5	sp-max-res-runs	1.310	[1.060, 1.579]
6	sp-restart-inc	1.501	[1.083, 1.971]
7	sp-variable-decay	1.404	[1.137, 1.683]
8	sp-clause-decay	1.397	[1.134, 1.667]
9	sp-rand-var-dec-freq	11.892	[1.641, 22.305]
10	sp-learned-size-factor	11.934	[1.709, 22.464]
11	sp-rand-phase-dec-freq	1.1474	[0.906, 1.412]
12	sp-use-pure-literal-rule	1.504	[1.101, 1.937]
13	sp-clause-activity-inc	1.455	[1.072, 1.867]
14	sp-clause-del-heur	1.481	[1.090, 1.913]
15	sp-update-dec-queue	1.488	[1.087, 1.920]
16	sp-learned-clause-sort-heur	11.327	[1.173, 21.641]
17	sp-res-cutoff-cls	11.390	[1.239, 21.752]
18	sp-rand-var-dec-scaling	93.552	[54.900,132.696]
19	sp-first-restart	2.256	[1.499, 3.113]
second-best	-	2.256	[1.499, 3.113]

Table 7 Parameters chosen and resulting test set performance for ablation analysis accelerated by racing (maximum of 200 rounds) for CPLEX on CORLAT test set, using the best ParamILS configuration as the source and the second-best ParamILS configuration as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance	Test bootstrap CI [10%, 90%]
best	-	5.411	[4.923, 5.922]
1	preprocessing_repeatpresolve	5.117	[4.615, 5.641]
2	mip_cuts_gomory	5.027	[4.541, 5.537]
3	simplex_limits_singularity	5.014	[4.542, 5.525]
4	simplex_pgradient	4.983	[4.495, 5.492]
5	preprocessing_boundstrength	4.989	[4.498, 5.483]
6	mip_limits_cutsfactor	5.026	[4.517, 5.563]
7	mip_strategy_dive	4.749	[4.285, 5.224]
8	barrier_crossover	4.753	[4.293, 5.246]
9	lpmethod	4.765	[4.305, 5.259]
10	mip_cuts_pathcut	4.751	[4.295, 5.228]
11	preprocessing_symmetry	4.760	[4.296, 5.241]
12	simplex_pricing	4.760	[4.298, 5.239]
13	barrier_limits_growth	4.767	[4.306, 5.252]
14	mip_cuts_zerohalfcut	4.747	[4.295, 5.227]
15	mip_cuts_cliques	4.751	[4.291, 5.237]
16	sifting_algorithm	4.745	[4.289, 5.228]
17	mip_strategy_presolvenode	4.780	[4.295, 5.274]
18	mip_strategy_variableselect	5.114	[4.628, 5.617]
19	mip_limits_cutpasses	5.658	[5.119, 6.241]
20	mip_strategy_probe	5.592	[5.033, 6.197]
21	simplex_tolerances_markowitz	5.485	[4.955, 6.032]
22	mip_cuts_flowcovers	6.286	[5.629, 6.963]
23	emphasis_mip	6.017	[5.417, 6.636]
24	mip_cuts_mircut	5.096	[4.543, 5.673]
25	mip_limits_aggforcut	4.823	[4.356, 5.319]
26	mip_cuts_covers	4.812	[4.343, 5.300]
27	mip_strategy_branch	5.036	[4.574, 5.512]
28	simplex_dgradient	5.024	[4.544, 4.532]
29	mip_strategy_heuristicfreq	5.932	[5.312, 6.575]
second-best	-	5.932	[5.312, 6.575]

Table 8 Parameters chosen and resulting test set performance for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the depots test set, using the best PARAMILS configuration as the source and the second-best PARAMILS configuration as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance	Test bootstrap CI [10%, 90%]
best	-	0.658	[0.622, 0.696]
1	extended_effects_evaluation	0.603	[0.569, 0.637]
2	static_noise	0.614	[0.579, 0.651]
3	numrestart	0.617	[0.581, 0.653]
4	evaluation_function	0.612	[0.577, 0.649]
5	triomemory	0.611	[0.576, 0.649]
6	dynoiseTabLen	0.616	[0.580, 0.654]
7	ri_list	0.617	[0.582, 0.654]
8	cri_update_iterations	0.621	[0.584, 0.659]
9	noise_incr	0.608	[0.574, 0.642]
10	dynoisecoefnum	0.627	[0.590, 0.666]
11	numtry	0.643	[0.605, 0.681]
12	weight_mutex_in_relaxed_plan	0.583	[0.554, 0.612]
13	walkplan	0.689	[0.650, 0.729]
14	tabu_length	0.585	[0.555, 0.616]
second-best	-	0.585	[0.555, 0.616]

Table 9 Parameters chosen and resulting test set performance for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the satellite test set, using the best ParamILS configuration as the source and the second-best ParamILS configuration as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance	Test bootstrap CI [10%, 90%]
best	-	5.783	[5.752, 5.814]
1	static_noise	5.293	[5.264, 5.322]
2	cri_update_iterations	5.293	[5.264, 5.323]
3	inc_re	5.239	[5.210, 5.270]
4	fast_best_action_evaluation	5.256	[5.226, 5.284]
5	evaluation_function	5.233	[5.204, 5.262]
6	evaluate_threated_supported_preconds...	5.234	[5.206, 5.263]
7	weight_mutex_in_relaxed_plan	5.213	[5.185, 5.242]
8	walkplan	5.257	[5.228, 5.287]
9	choose_min_numA_fact	5.211	[5.182, 5.239]
10	dynoiseTabLen	5.281	[5.251, 5.312]
11	mutex_and_additive_effects	5.269	[5.238, 5.300]
12	maxnoise	5.161	[5.132, 5.190]
13	not_supported_preconds_evaluation	5.220	[5.189, 5.251]
14	no_mutex_with_additive_effects	5.187	[5.157, 5.218]
15	dynoisecoefnum	5.214	[5.184, 5.245]
16	no_pruning	5.210	[5.180, 5.241]
17	noise_incr	5.189	[5.158, 5.218]
18	numtry	5.189	[5.160, 5.219]
19	no_insert_threated_act_in_neighb	5.145	[5.115, 5.174]
20	zero_num_A	5.144	[5.113, 5.174]
21	avoid_best_action_cycles	5.182	[5.152, 5.212]
22	noise	5.261	[5.230, 5.293]
23	twalkplan	5.170	[5.139, 5.201]
24	numrestart	5.198	[5.166, 5.229]
25	ri_list	5.117	[5.088, 5.148]
26	extended_effects_evaluation	5.157	[5.126, 5.187]
27	not_extended_unsupported_facts	5.356	[5.326, 5.387]
second-best	-	5.356	[5.326, 5.387]

Table 10 Parameters chosen and resulting test set performance for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the zenotravel test set, using the best ParamILS configuration as the source and the second-best ParamILS configuration as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance	Test bootstrap CI [10%, 90%]
best	-	2.092	[2.066, 2.119]
1	no_pruning	2.040	[2.013, 2.066]
2	criprecond	2.019	[1.993, 2.045]
3	choose_min_numA_fact	2.029	[2.003, 2.054]
4	inc_re	2.022	[1.996, 2.047]
5	remove_act_next_step	2.023	[1.996, 2.049]
6	nonuniform_random	2.016	[1.989, 2.042]
7	verifyAf	2.013	[1.987, 2.039]
8	numtry	2.009	[1.983, 2.034]
9	maxnoise	2.016	[1.991, 2.042]
10	numrestart	2.006	[1.981, 2.032]
11	no_insert_threated_act_in_neighb	2.019	[1.993, 2.045]
12	bestfirst	2.009	[1.982, 2.035]
13	dynoisecoefnum	2.013	[1.988, 2.038]
14	dynoiseTabLen	2.019	[1.985, 2.038]
15	verifyinit	2.025	[1.999, 2.050]
16	noise_incr	2.007	[1.981, 2.033]
17	cri_insertion_add_mutex	2.014	[1.988, 2.041]
18	evaluation_function	2.014	[1.988, 2.041]
19	extended_effects_evaluation	2.105	[2.078, 2.133]
20	h	2.119	[2.092, 2.146]
second-best	-	2.119	[2.092, 2.146]

Table 11 Parameters chosen and resulting test set performance for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the depots test set, using the depots configuration as the source and the satellite configuration as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance	Test bootstrap CI [10%, 90%]
depots	-	0.658	[0.622, 0.696]
1	extended_effects_evaluation	0.603	[0.569, 0.637]
2	criprecond	0.640	[0.602, 0.680]
3	ri_list	0.637	[0.599, 0.676]
4	verifyinit	0.642	[0.602, 0.683]
5	relaxed_examination	0.637	[0.598, 0.679]
6	fast_best_action_evaluation	0.593	[0.562, 0.625]
7	not_supported_preconds_evaluation	0.673	[0.628, 0.719]
8	inc_re	0.677	[0.635, 0.720]
9	zero_num_A	0.679	[0.638, 0.722]
10	cri_update_iterations	0.679	[0.636, 0.723]
11	maxnoise	0.678	[0.636, 0.720]
12	evaluation_function	0.674	[0.633, 0.716]
13	h	0.672	[0.631, 0.715]
14	numrestart	0.670	[0.628, 0.713]
15	improve_reachability	0.686	[0.643, 0.730]
16	mutex_and_additive_effects	0.626	[0.591, 0.663]
17	weight_mutex_in_relaxed_plan	0.617	[0.581, 0.653]
18	numtry	0.631	[0.593, 0.671]
19	evaluate_mutex_for_action_remotion	0.697	[0.650, 0.745]
20	not_extended_unsupported_facts	0.773	[0.721, 0.829]
21	no_mutex_with_additive_effects	0.866	[0.800, 0.936]
22	reset_extended_unsupported_facts	0.959	[0.882, 1.038]
23	noise	19.654	[13.579, 26.904]
24	cri_intermediate_levels	1005.004	[964.731, 1045.745]
25	hpar_cut_neighb	2643.840	[2615.603, 2670.889]
satellite	-	2643.840	[2615.603, 2670.889]

Table 12 Parameters chosen and resulting test set performance for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the satellite test set, using the satellite configuration as the source and the depots configuration as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance	Test bootstrap CI [10%, 90%]
depots	-	7.617	[7.576, 7.658]
1	hpar_cut_neighb	5.997	[5.964, 6.029]
2	not_extended_unsupported_facts	5.805	[5.773, 5.837]
3	reset_extended_unsupported_facts	5.710	[5.680, 5.741]
4	criprecond	5.340	[5.309, 5.369]
5	extended_effects_evaluation	5.290	[5.261, 5.318]
6	evaluate_mutex_for_action_remotion	5.559	[5.526, 5.592]
7	relaxed_examination	5.345	[5.315, 5.375]
8	inc_re	5.313	[5.284, 5.342]
9	ri_list	5.532	[5.501, 5.564]
10	not_supported_preconds_evaluation	5.307	[5.278, 5.336]
11	cri_intermediate_levels	5.368	[5.337, 5.399]
12	numrestart	5.405	[5.374, 5.436]
13	zero_num_A	5.376	[5.347, 5.404]
14	lagrange	5.376	[5.347, 5.404]
15	evaluation_function	5.375	[5.345, 5.404]
16	h	5.675	[5.640, 5.710]
17	noise	5.602	[5.568, 5.636]
18	cri_update_iterations	5.708	[5.674, 5.743]
19	weight_mutex_in_relaxed_plan	5.515	[5.483, 5.547]
20	improve_reachability	5.462	[5.431, 5.492]
21	no_mutex_with_additive_effects	5.427	[5.396, 5.458]
22	maxnoise	5.473	[5.442, 5.504]
23	numtry	5.668	[5.635, 5.701]
24	fast_best_action_evaluation	5.640	[5.607, 5.673]
25	mutex_and_additive_effects	5.297	[5.268, 5.326]
26	verifyinit	5.783	[5.752, 5.814]
satellite	-	5.783	[5.752, 5.814]

Table 13 Parameters chosen and resulting test set performance for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the depots test set, using the depots configuration as the source and the zenotravel configuration as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance	Test bootstrap CI [10%, 90%]
depots	-	0.658	[0.622, 0.696]
1	extended_effects_evaluation	0.603	[0.569, 0.637]
2	no_insert_threated_act_in_neighb	0.586	[0.555, 0.617]
3	lagrange	0.595	[0.564, 0.626]
4	no_pruning	0.595	[0.564, 0.627]
5	choose_min_numA_fact	0.576	[0.546, 0.606]
6	evaluation_function	0.584	[0.554, 0.615]
7	verifyinit	0.595	[0.564, 0.627]
8	zero_num_A	0.590	[0.559, 0.622]
9	nonuniform_random	0.601	[0.568, 0.633]
10	h	0.596	[0.565, 0.629]
11	bestfirst	0.578	[0.549, 0.608]
12	ri_list	0.588	[0.556, 0.619]
13	maxnoise	0.590	[0.559, 0.621]
14	cri_insertion_add_mutex	0.582	[0.552, 0.613]
15	triomemory	0.591	[0.559, 0.624]
16	numrestart	0.587	[0.554, 0.620]
17	static_noise	0.635	[0.593, 0.680]
18	inc_re	0.538	[0.512, 0.566]
19	criprecond	0.543	[0.514, 0.573]
20	dynoiseTabLen	0.542	[0.513, 0.572]
21	noise_incr	0.542	[0.513, 0.572]
22	dynoisecoefnum	0.542	[0.513, 0.572]
23	twalkplan	2.022	[0.509, 3.547]
24	walkplan	2.009	[0.497, 3.540]
25	numtry	0.535	[0.508, 0.563]
26	verifyAf	0.524	[0.498, 0.551]
27	reset_extended_unsupported_facts	0.526	[0.504, 0.549]
28	fast_best_action_evaluation	8.115	[3.630, 12.618]
29	hpar_cut_neighb	190.062	[169.054, 211.049]
30	noise	1275.739	[1233.667, 1319.085]
31	cri_intermediate_levels	2767.166	[2744.532, 2789.852]
32	stop_remove_act	2807.531	[2786.519, 2828.410]
zenotravel	-	2807.531	[2786.519, 2828.410]

Table 14 Parameters chosen and resulting test set performance for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the zenotravel test set, using the zenotravel configuration as the source and the depots configuration as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance	Test bootstrap CI [10%, 90%]
depots	-	1028.118	[988.133, 1066.608]
1	ri_list	1021.959	[982.752, 1062.075]
2	cri_intermediate_levels	10.058	[9.961, 10.155]
3	hpar_cut_neighb	2.755	[2.729, 2.781]
4	criprecond	3.584	[3.512, 3.658]
5	twalkplan	2.163	[2.137, 2.190]
6	extended_effects_evaluation	2.130	[2.105, 2.156]
7	noise	2.061	[2.035, 2.086]
8	stop_remove_act	2.020	[1.994, 2.048]
9	verifyAf	2.019	[1.992, 2.045]
10	numrestart	2.026	[1.999, 2.053]
11	triomemory	2.044	[2.017, 2.072]
12	walkplan	2.046	[2.019, 2.073]
13	verifyinit	2.046	[2.020, 2.074]
14	bestfirst	2.031	[2.005, 2.058]
15	cri_insertion_add_mutex	2.060	[2.034, 2.088]
16	static_noise	2.054	[2.027, 2.079]
17	zero_num_A	2.056	[2.029, 2.083]
18	maxnoise	2.077	[2.050, 2.104]
19	dynoiseTabLen	2.055	[2.029, 2.083]
20	lagrange	2.055	[2.029, 2.083]
21	no_insert_threated_act_in_neighb	2.077	[2.051, 2.104]
22	inc_re	2.082	[2.055, 2.109]
23	h	2.086	[2.060, 2.113]
24	dynoisecoefnum	2.108	[2.081, 2.135]
25	numtry	2.116	[2.089, 2.144]
26	choose_min_numA_fact	2.067	[2.040, 2.094]
27	evaluation_function	2.077	[2.050, 2.104]
28	reset_extended_unsupported_facts	2.063	[2.037, 2.089]
29	fast_best_action_evaluation	2.115	[2.089, 2.141]
30	noise_incr	2.088	[2.062, 2.115]
31	nonuniform_random	2.040	[2.013, 2.066]
32	no_pruning	2.092	[2.066, 2.119]
zenotravel	-	2.092	[2.066, 2.119]

Table 15 Parameters chosen and resulting test set performance for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the satellite test set, using the satellite configuration as the source and the zenotravel configuration as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance	Test bootstrap CI [10%, 90%]
satellite	-	5.783	[5.752, 5.814]
1	static_noise	5.293	[5.264, 5.322]
2	dynoisecoefnum	5.293	[5.264, 5.322]
3	no_pruning	5.767	[5.734, 5.800]
4	improve_reachability	5.485	[5.455, 5.517]
5	twalkplan	5.402	[5.373, 5.433]
6	not_extended_unsupported_facts	5.708	[5.674, 5.743]
7	nonuniform_random	5.778	[5.742, 5.811]
8	verifyAf	5.526	[5.493, 5.559]
9	noise_incr	5.631	[5.598, 5.666]
10	not_supported_preconds_evaluation	5.613	[5.578, 5.647]
11	dynoiseTabLen	5.681	[5.649, 5.713]
12	bestfirst	5.616	[5.582, 5.648]
13	cri_update_iterations	5.223	[5.194, 5.252]
14	cri_insertion_add_mutex	5.502	[5.470, 5.534]
15	evaluate_mutex_for_action_remotion	5.306	[5.276, 5.335]
16	inc_re	5.293	[5.263, 5.323]
17	h	5.385	[5.356, 5.415]
18	numtry	5.471	[5.439, 5.503]
19	no_mutex_with_additive_effects	5.322	[5.293, 5.352]
20	mutex_and_additive_effects	5.275	[5.247, 5.304]
21	triomemory	5.232	[5.202, 5.261]
22	stop_remove_act	5.222	[5.192, 5.251]
23	relaxed_examination	5.238	[5.209, 5.266]
24	weight_mutex_in_relaxed_plan	5.297	[5.269, 5.327]
25	numrestart	5.293	[5.264, 5.324]
26	walkplan	5.312	[5.282, 5.342]
27	no_insert_threated_act_in_neighb	5.355	[5.325, 5.386]
28	choose_min_numA_fact	5.291	[5.262, 5.320]
29	maxnoise	5.357	[5.327, 5.387]
30	cri_intermediate_levels	7.952	[7.908, 7.996]
zenotravel	-	7.952	[7.908, 7.996]

Table 16 Parameters chosen and resulting test set performance for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the zenotravel test set, using the zenotravel configuration as the source and the satellite configuration as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance	Test bootstrap CI [10%, 90%]
satellite	-	2478.914	[2446.288, 2511.701]
1	cri_intermediate_levels	2.954	[2.869, 3.045]
2	stop_remove_act	2.142	[2.113, 2.171]
3	not_supported_preconds_evaluation	2.280	[2.234, 2.326]
4	walkplan	2.084	[2.057, 2.111]
5	not_extended_unsupported_facts	2.084	[2.057, 2.110]
6	numtry	2.096	[2.070, 2.123]
7	evaluate_mutex_for_action_remotion	2.121	[2.094, 2.148]
8	cri_update_iterations	2.062	[2.036, 2.088]
9	mutex_and_additive_effects	2.022	[1.996, 2.048]
10	relaxed_examination	2.021	[1.995, 2.048]
11	maxnoise	2.019	[1.993, 2.045]
12	cri_insertion_add_mutex	2.020	[1.994, 2.045]
13	weight_mutex_in_relaxed_plan	2.017	[1.991, 2.042]
14	numrestart	2.020	[1.994, 2.046]
15	no_mutex_with_additive_effects	2.031	[2.005, 2.057]
16	triomemory	2.038	[2.012, 2.064]
17	choose_min_numA_fact	2.035	[2.009, 2.061]
18	verifyAf	2.051	[2.025, 2.076]
19	improve_reachability	2.040	[2.014, 2.066]
20	inc_re	2.056	[2.030, 2.082]
21	h	2.097	[2.071, 2.124]
22	no_insert_threated_act_in_neighb	2.050	[2.024, 2.076]
23	twalkplan	2.086	[2.060, 2.113]
24	static_noise	2.066	[2.039, 2.093]
25	noise_incr	2.104	[2.077, 2.130]
26	bestfirst	2.088	[2.062, 2.115]
27	nonuniform_random	2.040	[2.013, 2.066]
28	no_pruning	2.092	[2.066, 2.119]
zenotravel	-	2.092	[2.066, 2.119]

Table 17 Parameters chosen and resulting test set performance for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the zenotravel test set, using the depots configuration as the source and the satellite configuration as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance	Test bootstrap CI [10%, 90%]
depots	-	1028.131	[989.502, 1067.367]
1	hpar_cut_neighb	605.424	[572.784, 638.556]
2	extended_effects_evaluation	35.646	[32.356, 39.224]
3	not_supported_preconds_evaluation	26.560	[25.672, 27.449]
4	fast_best_action_evaluation	22.093	[20.227, 24.195]
5	reset_extended_unsupported_facts	18.780	[16.991, 20.897]
6	no_mutex_with_additive_effects	17.173	[16.531, 17.829]
7	not_extended_unsupported_facts	18.690	[16.854, 20.727]
8	improve_reachability	18.594	[16.765, 20.683]
9	verifyinit	18.628	[16.778, 20.698]
10	numrestart	18.599	[16.747, 20.716]
11	maxnoise	18.579	[16.739, 20.667]
12	cri_update_iterations	18.550	[16.721, 20.659]
13	evaluation_function	18.511	[16.658, 20.569]
14	h	18.508	[16.681, 20.604]
15	ri_list	18.544	[16.707, 20.620]
16	relaxed_examination	18.502	[16.666, 20.616]
17	zero_num_A	18.492	[16.672, 20.612]
18	inc_re	17.706	[15.913, 19.798]
19	numtry	21.087	[17.920, 24.494]
20	criprecond	19.495	[16.774, 22.432]
21	evaluate_mutex_for_action_remotion	26.080	[22.802, 29.539]
22	mutex_and_additive_effects	28.061	[24.765, 31.568]
23	weight_mutex_in_relaxed_plan	31.302	[28.051, 34.888]
24	noise	38.765	[33.719, 44.109]
25	cri_intermediate_levels	2478.914	[2446.639, 2511.390]
satellite	-	2478.914	[2446.639, 2511.390]

Table 18 Parameters chosen and resulting test set performance for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the satellite test set, using the depots configuration as the source and the zenotravel configuration as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance	Test bootstrap CI [10%, 90%]
depots	-	7.620	[7.580, 7.661]
1	hpar_cut_neighb	5.997	[5.964, 6.029]
2	criprecond	5.719	[5.688, 5.750]
3	walkplan	5.552	[5.520, 5.583]
4	lagrange	5.733	[5.702, 5.763]
5	verifyAf	5.706	[5.675, 5.739]
6	static_noise	5.567	[5.536, 5.599]
7	dynoisecoefnum	5.567	[5.536, 5.599]
8	noise	5.470	[5.438, 5.501]
9	maxnoise	5.574	[5.542, 5.606]
10	inc_re	5.584	[5.553, 5.615]
11	triomemory	5.486	[5.454, 5.519]
12	no_pruning	5.363	[5.331, 5.395]
13	numtry	5.304	[5.274, 5.335]
14	numrestart	5.245	[5.215, 5.274]
15	reset_extended_unsupported_facts	5.311	[5.280, 5.343]
16	noise_incr	5.417	[5.384, 5.450]
17	verifyinit	5.382	[5.349, 5.415]
18	twalkplan	5.335	[5.302, 5.367]
19	dynoiseTabLen	5.310	[5.278, 5.342]
20	no_insert_threated_act_in_neighb	5.460	[5.428, 5.492]
21	evaluation_function	5.638	[5.606, 5.672]
22	fast_best_action_evaluation	5.441	[5.410, 5.474]
23	h	5.460	[5.428, 5.493]
24	extended_effects_evaluation	5.496	[5.463, 5.529]
25	bestfirst	5.421	[5.388, 5.452]
26	zero_num_A	5.325	[5.294, 5.356]
27	nonuniform_random	5.454	[5.420, 5.486]
28	ri_list	5.399	[5.366, 5.432]
29	cri_insertion_add_mutex	5.310	[5.278, 5.341]
30	stop_remove_act	5.409	[5.375, 5.442]
31	choose_min_numA_fact	5.568	[5.536, 5.599]
32	cri_intermediate_levels	7.951	[7.907, 7.995]
zenotravel	-	7.951	[7.907, 7.995]

Table 19 Parameters chosen and resulting test set performance for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the depots test set, using the satellite configuration as the source and the zenotravel configuration as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance	Test bootstrap CI [10%, 90%]
satellite	-	2643.840	[2615.603, 2670.889]
1	walkplan	1386.764	[1343.744, 1429.855]
2	verifyAf	1306.732	[1263.422, 1349.898]
3	numtry	1309.948	[1266.984, 1351.802]
4	static_noise	1246.085	[1204.583, 1287.956]
5	dynoiseTabLen	1174.005	[1132.051, 1215.895]
6	cri_insertion_add_mutex	1173.921	[1132.164, 1215.830]
7	noise_incr	1173.913	[1131.960, 1215.687]
8	not_supported_preconds_evaluation	1181.307	[1139.149, 1223.153]
9	nonuniform_random	1181.321	[1139.506, 1223.338]
10	relaxed_examination	1181.324	[1139.290, 1223.039]
11	no_pruning	1181.252	[1138.278, 1223.058]
12	triomemory	1181.313	[1139.235, 1223.602]
13	improve_reachability	1181.309	[1139.408, 1222.830]
14	dynoisecoefnum	1181.309	[1139.577, 1223.528]
15	maxnoise	1181.284	[1140.522, 1224.623]
16	h	1181.300	[1140.564, 1223.085]
17	bestfirst	1181.288	[1139.333, 1223.190]
18	cri_update_iterations	1181.284	[1139.471, 1223.320]
19	choose_min_numA_fact	1181.288	[1139.547, 1223.467]
20	evaluate_mutex_for_action_remotion	1122.883	[1081.979, 1164.738]
21	not_extended_unsupported_facts	1070.618	[1029.043, 1111.062]
22	no_mutex_with_additive_effects	1074.977	[1033.327, 1116.917]
23	weight_mutex_in_relaxed_plan	1116.059	[1075.131, 1158.243]
24	mutex_and_additive_effects	1098.086	[1056.794, 1138.603]
25	no_insert_threated_act_in_neighb	1181.602	[1140.940, 1223.464]
26	numrestart	1701.370	[1659.336, 1743.315]
27	inc_re	1843.723	[1801.749, 1887.203]
28	stop_remove_act	2148.242	[2109.262, 2187.206]
29	cri_intermediate_levels	2575.323	[2545.877, 2605.028]
30	twalkplan	2807.531	[2786.519, 2828.410]
zenotravel	-	2807.531	[2786.519, 2828.410]

Table 20 Parameters chosen and resulting PAR10 performance on both the training and test sets for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the depots training set, using our extended approach with support for conditional parameters. Ablation was performed using the LPG default configuration as the source, and the configuration produced by PARAMILS as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance		Test bootstrap CI
		Training	Test	[10%, 90%]
default	-	39.635	34.874	[27.506, 42.563]
1	cri_intermediate_levels	1.577	1.470	[1.375, 1.569]
2	vicinato, hpar_cut_neighb	0.920	0.910	[0.870, 0.952]
3	triomemory	0.818	0.817	[0.772, 0.863]
4	donot_try_suspected_actions	0.790	0.780	[0.738, 0.824]
5	noise	0.729	0.689	[0.655, 0.725]
6	walkplan	0.695	0.626	[0.598, 0.656]
7	static_noise	0.686	0.652	[0.619, 0.685]
8	numtry	0.620	0.593	[0.563, 0.625]
9	extended_effects_evaluation	0.687	0.659	[0.625, 0.694]
10	maxnoise	0.693	0.662	[0.628, 0.698]
11	weight_mutex_in_relaxed_plan	0.640	0.645	[0.610, 0.680]
12	inc_re	0.648	0.646	[0.612, 0.681]
13	ri_list	0.648	0.640	[0.607, 0.675]
14	numrestart	0.646	0.638	[0.605, 0.672]
configured	-	0.646	0.638	[0.605, 0.672]

Table 21 Parameters chosen and resulting PAR10 performance on both the training and test sets for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the satellite training set, using our extended approach with support for conditional parameters. Ablation was performed using the LPG default configuration as the source, and the configuration produced by PARAMILS as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance		Test bootstrap CI
		Training	Test	[10%, 90%]
default	-	18.890	18.728	[17.192, 20.312]
1	cri_intermediate_levels	7.014	7.040	[7.005, 7.075]
2	vicinato, hpar_cut_neighb	5.739	5.735	[5.706, 5.764]
3	criprecond	5.354	5.343	[5.316, 5.370]
4	evaluation_function	5.333	5.344	[5.317, 5.370]
5	noise	5.307	5.323	[5.295, 5.351]
6	not_extended_unsupported_facts	5.310	5.323	[5.297, 5.350]
7	not_supported_preconds_evaluation	5.297	5.311	[5.284, 5.337]
8	static_noise	5.312	5.326	[5.299, 5.353]
9	improve_reachability	5.302	5.327	[5.300, 5.354]
10	reset_extended_unsupported_facts	5.308	5.315	[5.288, 5.342]
11	fast_best_action_evaluation	5.337	5.320	[5.293, 5.346]
12	maxnoise	5.316	5.337	[5.309, 5.364]
13	evaluate_mutex_for_action_remotion	5.289	5.312	[5.286, 5.339]
14	donot_try_suspected_actions	5.298	5.306	[5.279, 5.333]
15	verifyinit	5.302	5.304	[5.276, 5.331]
16	triomemory	5.267	5.306	[5.280, 5.332]
17	zero_num_A	5.281	5.299	[5.273, 5.325]
18	mutex_and_additive_effects	5.272	5.292	[5.266, 5.318]
19	cri_update_iterations	5.260	5.293	[5.266, 5.320]
20	walkplan	5.260	5.310	[5.282, 5.339]
21	numtry	5.274	5.317	[5.291, 5.344]
22	h	5.264	5.300	[5.273, 5.327]
23	numrestart	5.277	5.324	[5.297, 5.351]
24	relaxed_examination	5.279	5.305	[5.279, 5.332]
25	no_mutex_with_additive_effects	5.275	5.302	[5.276, 5.329]
26	inc_re	5.310	5.273	[5.246, 5.301]
configured	-	5.310	5.273	[5.246, 5.301]

Table 22 Parameters chosen and resulting PAR10 performance on both the training and test sets for ablation analysis accelerated by racing (maximum of 200 rounds) for LPG on the zenotravel training set, using our extended approach with support for conditional parameters. Ablation was performed using the LPG default configuration as the source, and the configuration produced by PARAMILS as the target. The "Test bootstrap CI" column indicates the [10%, 90%] confidence interval of PAR10 performance, using 10 000 bootstrap samples of our test set data for each configuration.

Round	Parameter	Performance		Test bootstrap CI
		Training	Test	[10%, 90%]
default	-	27.326	27.293	[25.630, 29.146]
1	vicinato, hpar_cut_neighb	4.847	4.753	[4.712, 4.795]
2	triomemory	3.016	2.981	[2.950, 3.012]
3	criprecond	2.459	2.413	[2.385, 2.440]
4	fast_best_action_evaluation	2.213	2.157	[2.130, 2.182]
5	stop_remove_act	2.126	2.078	[2.049, 2.107]
6	noise	1.974	1.933	[1.907, 1.959]
7	bestfirst	1.977	1.916	[1.890, 1.941]
8	donot_try_suspected_actions	1.968	1.912	[1.886, 1.938]
9	weight_mutex_in_relaxed_plan	1.961	1.904	[1.878, 1.930]
10	cri_insertion_add_mutex	1.961	1.911	[1.885, 1.937]
11	reset_extended_unsupported_facts	1.970	1.916	[1.890, 1.942]
12	numtry	1.972	1.915	[1.889, 1.940]
13	no_pruning	1.973	1.914	[1.888, 1.939]
14	no_insert_threated_act_in_neighb	1.978	1.920	[1.893, 1.945]
15	numrestart	1.976	1.915	[1.890, 1.940]
16	maxnoise	1.971	1.914	[1.888, 1.939]
17	verifyinit	1.972	1.916	[1.890, 1.942]
18	verifyAf	1.964	1.916	[1.891, 1.942]
19	twalkplan	1.968	1.911	[1.886, 1.937]
20	noise_incr	1.969	1.918	[1.892, 1.944]
21	inc_re	1.969	1.918	[1.892, 1.944]
22	zero_num_A	1.971	1.921	[1.895, 1.946]
23	choose_min_numA_fact	1.968	1.907	[1.881, 1.932]
24	evaluation_function	1.969	1.909	[1.884, 1.934]
25	nonuniform_random	1.974	1.917	[1.890, 1.942]
26	dynoiseTabLen	1.970	1.933	[1.907, 1.957]
27	h	1.955	1.939	[1.913, 1.965]
configured	-	1.955	1.939	[1.913, 1.965]

View publication stats