CENTRO DE CIENCIAS EXATAS E TECNOLOGICAS - CCE

DEPARTAMENTO DE INFORMATICA

REPORT:

COMPARISON BETWEEN SS AND IDA*

Marvin Abisrror Zarate
MSc Student in Computer Science

Levi Henrique Santana de Lelis
(Advisor)

Santiago Franco
(Co-Advisor)

VIÇOSA - MINAS GERAIS
JULY - 2015

Contents

1 Introduction

The new approach for selecting a subset of heuristics functions for domain-independent planning has two main objectives: First, make a selection of heuristics from a large set of heuristics with the goal of reducing the running time of a search algorithm employing the subset functions. Second, find out if the prediction of Stratified Sampling (SS) might be helpful in selecting a subset of heuristics to guide the A* search.

In order to achieve the first objective we present The Greedy Algorithm, which provides a good approximation to the optimal solution of the NP-hard optimization problem [Krause and Golovin, 2012].

In order to achieve the second objective we use the *relative unsigned error* to probe the accuracy of the predictions of SS with respect to IDA*. We know that SS does not make even reasonable predictions for the number of nodes expanded by A*. Nevertheless, even though SS produces poor predictions for the number of nodes expanded by A*, we would like to verify whether these predictions can be helpful in selecting a subset of heuristics to guide the A* search.

This report have three sections, the first section is the introduction, the second is the experiment 1 which contains four main tables showing the results of the *relative unsigned error*. and the last section in the conclusion.

2 Comparison of SS with IDA*

Stratification Sampling is an algorithm that estimates the number of nodes expanded performed by heuristic search algorithm seeking solutions in state space. We apply SS to predict the number of nodes expanded by IDA* in a given f-layer when using a consistent heuristics.

We first ran IDA* for Fast-Downward benchmark for optimal domains. Our evaluation metric is coverage, *i.e.,* number of problems solved within 24 hours time limit. We note that in 24 hours non all the instances for a specific domain using a consistent heuristic can be solved. Afterwards, run SS using as a threshold the f-layer for each instance of each domain, this process is executed using different number of probes *i.e.,* 1, 10, 100, 1000, and 5000.

In our experiment 1, prediction accuracy is measured in terms of the *Relative Unsigned Error* (ss-err), which is calculated as:

$$\frac{\sum_{s \in PI} \frac{Pred(s,d) - R(s,d)}{R(s,d)}}{|PI|}$$

Where *PI* is the set of problem instances, *Pred(s, d)* and *R(s, d)* are the predicted and actual number of nodes expanded by IDA* for start state s and cost bound d. A perfect score according to this measure is 0.000.

The heuristics used for this experiment 1 were: hmax, ipdb, lmcut, and merge_and_shrink. There are 4 tables, each table shows the results running IDA* and SS using one consistent heuristic. The first column represent the optimal domains for Fast-Downward benchmark. The remaining 10 columns shows the 5 different probes *i.e.,* 1, 10, 100, 1000, and 5000. Each probe has two columns which represent the ss-err and the ss-time. The last two columns are the information for IDA* which represent the average value of the number of nodes expanded and the average time respectively. The text "—" means that IDA* could not solve the problems, consequently there are not results for SS.

Table 1: Experiment 1 - Comparison using hmax heuristic

	hmax											
	1		10		100		1000		5000			
Domain	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ida*	ida-time
barman-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
blocks	2.648	0.003	1.113	0.178	0.498	2.266	0.159	23.300	0.087	116.332	156474000.000	10770.900
elevators-opt08-strips	5.530	7.320	4.823	75.700	1.791	739.500	1.777	7571.860	1.551	37949.900	1604640.000	4358.740
elevators-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
floortile-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
nomystery-opt11-strips	9.204	0.043	2.465	0.363	0.735	3.707	0.369	36.847	0.130	183.897	2911010.000	289.453
openstacks-opt08-adl	—	—	—	—	—	—	—	—	—	—	—	—
openstacks-opt08-strips	0.192	0.668	0.141	6.636	0.128	67.228	0.141	679.132	0.146	3365.820	361346.000	786.512
openstacks-opt11-strips	0.000	0.860	0.000	8.300	0.000	80.400	0.000	798.200	0.000	4196.760	195.000	2.560
parcprinter-opt11-strips	0.000	0.003	0.000	0.035	0.000	0.392	0.000	3.344	0.000	17.286	1.000	0.000
parking-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
pegsol-opt11-strips	0.285	0.011	0.107	0.084	0.028	0.852	0.027	7.933	0.007	39.331	151280.000	10.998
scanalyzer-opt11-strips	0.961	0.080	0.177	0.120	0.118	0.880	0.026	8.080	0.019	39.920	101.000	0.160
sokoban-opt08-strips	0.114	0.035	0.308	0.351	0.286	3.518	0.023	30.990	0.015	155.113	5545.650	1.813
sokoban-opt11-strips	1.795	0.035	0.138	0.329	0.123	3.323	0.025	33.011	0.015	162.382	8245.920	2.622
tidybot-opt11-strips	3.754	5.727	1.582	29.967	0.608	287.373	0.269	2843.170	0.186	15769.200	2716460.000	17900.100
transport-opt08-strips	0.622	0.000	0.222	0.000	0.025	0.040	0.018	0.800	0.007	4.080	38.000	0.000
transport-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
visitall-opt11-strips	0.445	0.000	0.190	0.003	0.073	0.277	0.017	3.200	0.016	16.246	5363080.000	96.117
woodworking-opt08-strips	7.063	0.095	5.483	1.245	1.702	12.900	0.529	129.285	0.345	646.345	5190350.000	1023.720
woodworking-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—

Table 2: ss-err variation

	1		10		100		1000		5000			
Domain	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ida*	ida-time
openstacks-opt08-strips	**0.192**	0.668	**0.141**	6.636	**0.128**	67.228	**0.141**	679.132	**0.146**	3365.820	361346.000	786.512

In the Table 1 we can see that there are seven domains which IDA* could not solve any instance in 24 hours. The ss-err decrease for each domain according the number of probes increases. The domains that have the perfect score are: openstacks-opt11-strips, parcprinter-opt11-strips.

In the Table 2 the ss-err with 1 probe is **0.192**, afterwards it descrease with 10 probes to **0.141**, afterwards it decrease to **0.128**, but with 1000 probes instead to continue decreasing the opposite happens, it increase to **0.141** and continue increasing using 5000 probes to **0.146**. This kind of results might be due the fact that SS can consider more predictions and that some of them increase the result of the ss-err.

Table 3: Experiment 1 - Comparison using ipdb heuristic

	ipdb											
	1		10		100		1000		5000			
Domain	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ida*	ida-time
barman-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
blocks	4.789	0.002	3.543	0.013	1.678	0.134	0.742	1.366	0.485	6.909	2096930000.000	10018.200
elevators-opt08-strips	8.909	0.835	5.724	8.550	4.030	85.280	1.583	858.460	1.183	4287.430	95651000.000	20223.900
elevators-opt11-strips	13.495	1.130	9.043	11.380	3.898	117.380	2.846	1178.090	1.771	5904.640	154128000.000	36205.900
floortile-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
nomystery-opt11-strips	1429.530	0.003	1.759	0.011	1.065	0.114	0.475	1.103	0.167	5.536	360878000.000	2514.160
openstacks-opt08-adl	—	—	—	—	—	—	—	—	—	—	—	—
openstacks-opt08-strips	0.376	0.820	0.301	7.631	0.279	75.666	0.293	772.126	0.290	4522.480	2043600.000	15633.000
openstacks-opt11-strips	0.316	1.827	0.449	16.107	0.457	172.507	0.442	1678.330	0.437	9491.130	4166230.000	35539.000
parcprinter-opt11-strips	0.040	0.000	0.011	0.006	0.006	0.071	0.000	0.735	0.077	3.603	3065.690	337.491
parking-opt11-strips	3.584	0.028	1.488	0.200	0.399	2.116	0.135	20.748	0.288	103.608	306957000.000	7448.660
pegsol-opt11-strips	1.515	0.047	0.662	0.387	0.342	3.973	0.327	39.435	0.376	190.308	54945.300	1181.480
scanalyzer-opt11-strips	1.353	0.003	49.072	0.097	0.167	0.857	0.342	8.326	0.036	42.423	1071520000.000	8163.320
sokoban-opt08-strips	0.112	0.006	0.026	0.035	0.005	0.372	0.004	3.654	0.003	18.271	280429.000	9.525
sokoban-opt11-strips	0.073	0.007	0.001	0.049	0.004	0.482	0.002	4.869	0.001	24.387	404338.000	13.458
tidybot-opt11-strips	1.759	0.080	0.637	0.853	0.416	8.400	0.096	84.127	0.036	420.853	1131220.000	355.400
transport-opt08-strips	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.040	0.000	0.260	15.000	0.120
transport-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
visitall-opt11-strips	1.136	0.002	0.784	0.015	0.234	0.080	0.075	0.840	0.055	3.989	7604630.000	31.693
woodworking-opt08-strips	2.541	0.037	0.950	0.287	0.384	2.817	0.092	27.843	0.054	138.885	93740200.000	6484.690
woodworking-opt11-strips	3.093	0.060	1.808	0.573	0.354	5.913	0.248	59.680	0.116	295.787	249863000.000	16954.900

Table 4: ss-err variation

	1		10		100		1000		5000			
Domain	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ida*	ida-time
parcprinter-opt11-strips	**0.040**	0.000	**0.011**	0.006	**0.006**	0.071	**0.000**	0.735	**0.077**	3.603	3065.690	337.491

In Table 3 we can see that there are four domains which IDA* could not solve any instance in 24 hours. The ss-err decrease for most of the domains domain according the number of probes increases. There are few exception for example the domain parcrinter-opt11-strips that have the following result:

In the Table 4 the perfect score is found in the probe 1000. In the probe 5000 the ss-err change to 0.077. The true is that there are not better number than the perfect score. When SS is run using 5000 probes is possible that exists quite changes in the results that increase the result of the ss-err.

Table 5: Experiment 1 - Comparison using lmcut heuristic

	lmcut											
	1		10		100		1000		5000			
Domain	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ida*	ida-time
barman-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
blocks	3.266	0.164	2.120	1.667	0.708	15.701	0.290	157.335	0.174	801.386	9943800.000	9361.030
elevators-opt08-strips	5.720	70.105	3.457	691.590	0.911	6210.870	0.299	61445.500	0.242	323240.000	139417.000	1458.650
elevators-opt11-strips	6.570	117.520	4.427	1190.320	1.717	10650.500	0.642	107014.000	0.310	577514.000	220608.000	2573.950
floortile-opt11-strips	12.185	1.340	3.174	14.740	2.227	132.460	0.450	1311.420	0.256	7195.940	25413400.000	21021.000
nomystery-opt11-strips	4.515	1.262	1.490	14.763	0.675	124.755	0.204	1262.320	0.144	6858.100	647051.000	5214.300
openstacks-opt08-adl	—	—	—	—	—	—	—	—	—	—	—	—
openstacks-opt08-strips	0.092	0.544	0.157	5.972	0.148	58.472	0.140	576.712	0.147	3911.320	361346.000	677.368
openstacks-opt11-strips	0.000	1.060	0.000	10.300	0.000	66.780	0.000	672.900	0.000	3698.520	195.000	2.320
parcprinter-opt11-strips	0.005	0.038	0.006	0.362	0.000	3.558	0.000	36.046	0.000	182.945	3064.850	0.045
parking-opt11-strips	1.869	7.060	0.992	75.100	0.210	732.120	0.211	7424.720	0.177	38761.900	286812.000	4761.780
pegsol-opt11-strips	21.763	0.098	0.373	0.868	0.240	7.798	0.213	74.579	0.159	393.330	1130870.000	630.318
scanalyzer-opt11-strips	1.112	13.417	0.383	128.737	0.094	1176.560	0.039	11800.200	0.024	76516.400	7211710.000	12482.300
sokoban-opt08-strips	0.028	0.527	0.038	5.053	0.003	50.950	0.003	507.744	0.001	2915.320	232738.000	287.884
sokoban-opt11-strips	0.038	0.678	0.005	6.507	0.004	63.178	0.002	633.602	0.001	3260.390	295998.000	368.898
tidybot-opt11-strips	4.918	19.740	1.818	186.110	0.685	1788.940	0.280	17720.700	0.159	95541.000	1162760.000	17561.200
transport-opt08-strips	4.542	0.360	1.012	3.993	0.392	38.387	0.125	384.567	0.049	2069.430	712491.000	778.980
transport-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
visitall-opt11-strips	1.441	0.022	3.187	0.456	0.175	4.473	0.066	46.993	0.048	237.667	595466.000	91.538
woodworking-opt08-strips	1.123	0.211	0.266	1.854	0.148	18.497	0.019	188.931	0.029	1051.260	6048240.000	1859.450
woodworking-opt11-strips	0.789	0.320	0.131	3.490	0.228	32.830	0.063	322.000	0.013	1654.180	21020800.000	6462.270

In the Table 5 we can see that there are three domains which IDA* could not solve any instance in 24 hours. The ss-err decrease for most of the domains according the number of probes increases, except for openstacks-opt08-strips, parcprinter-opt11-strips, sokoban-opt08-strips, visitall-opt11-strips. Nevertheless, the decrease is not a pattern, it means that only occur between two consecutive probes and the difference is minimal. The domains that have the perfect score are: openstacks-opt11-strips and parcprinter-opt11-strips.

Table 6: Experiment 1 - Comparison using merge_and_shrink heuristic

	merge_and_shrink											
	1		10		100		1000		5000			
Domain	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ida*	ida-time
barman-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
blocks	2.163	13.074	3.346	13.444	0.482	14.151	0.205	14.769	0.256	17.972	5807970.000	75.288
elevators-opt08-strips	2.479	4.690	1.968	7.370	1.293	34.820	0.970	292.250	0.924	1440.010	69178500.000	14434.600
elevators-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
floortile-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
nomystery-opt11-strips	0.766	0.861	0.453	0.866	0.289	1.027	0.194	1.374	0.212	3.366	800639000.000	3766.920
openstacks-opt08-adl	—	—	—	—	—	—	—	—	—	—	—	—
openstacks-opt08-strips	0.666	37.090	0.518	26.655	0.182	54.095	0.175	180.090	0.181	838.820	451632.000	358.620
openstacks-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
parcprinter-opt11-strips	0.055	16.812	0.008	17.260	0.005	16.925	0.001	17.323	0.000	17.846	3064.850	16.872
parking-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
pegsol-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
scanalyzer-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
sokoban-opt08-strips	0.039	59.195	0.011	56.500	0.000	75.720	0.000	67.545	0.000	116.690	4135.000	57.275
sokoban-opt11-strips	0.000	235.760	0.000	267.320	0.000	371.140	0.000	318.020	0.000	458.180	5.000	223.340
tidybot-opt11-strips	0.000	1211.100	1.000	0.000	0.000	1206.820	0.000	1592.380	0.000	1326.720	65.000	1823.720
transport-opt08-strips	0.139	0.016	0.025	0.020	0.004	0.044	0.003	0.172	0.004	0.804	1391.800	0.024
transport-opt11-strips	—	—	—	—	—	—	—	—	—	—	—	—
visitall-opt11-strips	0.533	0.818	0.416	0.902	0.180	0.907	0.036	1.058	0.019	2.318	2613370.000	9.673
woodworking-opt08-strips	0.707	9.966	0.400	11.906	0.137	11.640	0.041	17.886	0.025	47.763	10699200.000	303.166
woodworking-opt11-strips	1.686	32.360	1.761	35.480	0.221	43.670	0.062	58.410	0.690	113.240	36373700.000	1057.370

In the Table 6 we can see that there are nine domains which IDA* could not solve any instance in 24 hours. The ss-err decrease for each domain according the number of probes increases. The domains that have the perfect score are: parcprinter-opt11-strips, sokoban-opt08-strips, sokoban-opt11-strips, and tidybot-opt11-strips.

3 Conclusions

The Table 6 using merge_and_shrink has more unsolved domains and one of the reason why this happen is because it throws run out memory for most of the problems. However, is the same Table 6 which contains more perfect scores and the reason could be that merge_and_shrink generates more precise pattern database when it initializes and give us better heuristics to guide the search. The Table 5 shows that the heuristic lmcut solved most of the domains in 24 hours, which implies that lmcut give us better heuristics too. The domains that could not be solved by all the consistent heuristic were openstacks-opt08-adl, barman-opt11-strips, and transport-opt11-strips. The problem of openstacks-opt08-adl is that do not support axioms and the other domains contains difficult instances that were not possible to solve any instance within 24 hours.

Generally, the ss-err decreases for the most part of the domains when the number of probes increases. I think SS gives us good predictions for IDA*.

Table 7: Experiment 1 - Comparison using hmax heuristic

	hmax											
	1		10		100		1000		5000			
Domain	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ss-err	ss-t	ida*	ida-time
barman-opt11	—	—	—	—	—	—	—	—	—	—	—	—
blocks	2.65	0.00	1.11	0.18	0.50	2.27	0.16	23.30	0.09	116.33	156474000.00	10770.90
elevators-opt08	5.53	7.32	4.82	75.70	1.79	739.50	1.78	7571.86	1.55	37949.90	1604640.00	4358.74
elevators-opt11	—	—	—	—	—	—	—	—	—	—	—	—
floortile-opt11	—	—	—	—	—	—	—	—	—	—	—	—
nomystery-opt11	9.20	0.04	2.47	0.36	0.74	3.71	0.37	36.85	0.13	183.90	2911010.00	289.45
openstacks-adl	—	—	—	—	—	—	—	—	—	—	—	—
openstacks-opt08	0.19	0.67	0.14	6.64	0.13	67.23	0.14	679.13	0.15	3365.82	361346.00	786.51
openstacks-opt11	0.00	0.86	0.00	8.30	0.00	80.40	0.00	798.20	0.00	4196.76	195.00	2.56
parcprinter-opt11	0.00	0.00	0.00	0.04	0.00	0.39	0.00	3.34	0.00	17.29	1.00	0.00
parking-opt11	—	—	—	—	—	—	—	—	—	—	—	—
pegsol-opt11	0.29	0.01	0.11	0.08	0.03	0.85	0.03	7.93	0.01	39.33	151280.00	11.00
scanalyzer-opt11	0.96	0.08	0.18	0.12	0.12	0.88	0.03	8.08	0.02	39.92	101.00	0.16
sokoban-opt08	0.11	0.04	0.31	0.35	0.29	3.52	0.02	30.99	0.02	155.11	5545.65	1.81
sokoban-opt11	1.80	0.04	0.14	0.33	0.12	3.32	0.02	33.01	0.01	162.38	8245.92	2.62
tidybot-opt11	3.75	5.73	1.58	29.97	0.61	287.37	0.27	2843.17	0.19	15769.20	2716460.00	17900.10
transport-opt08	0.62	0.00	0.22	0.00	0.02	0.04	0.02	0.80	0.01	4.08	38.00	0.00
transport-opt11	—	—	—	—	—	—	—	—	—	—	—	—
visitall-opt11	0.45	0.00	0.19	0.00	0.07	0.28	0.02	3.20	0.02	16.25	5363080.00	96.12
woodworking-opt08	7.06	0.10	5.48	1.25	1.70	12.90	0.53	129.28	0.35	646.35	5190350.00	1023.72
woodworking-opt08	—	—	—	—	—	—	—	—	—	—	—	—

References

Andreas Krause and Daniel Golovin. Submodular function maximization. *Tractability: Practical Approaches to Hard Problems*, 3:19, 2012.