

Estimation of parameter importance and interaction with Random Forests

Leslie Pérez Cáceres

In this document we are testing how to use Random Forest to assess importance and interactions of parameters based on the data gathered by irace.

For the analysis below we use as example ACOTSP (long):

- 20 secs cut off time (long)
- 11 parameters
- 400 instances sizes (1000,1500,2000,2500,3000,3500,4000,4500)
- 5000 experiments configuration budget

1. Importance by iterations

Iteration 1:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
2	alpha	1.816667	7950	579.02164	239019.38	300	34	0.0000000
9	localsearch	1.840000	2384	816.10954	338744.58	300	67	1.0000000
8	instance	2.180000	15346	589.31632	327768.74	300	0	0.0000000
10	nnls	2.343333	5537	641.34674	287970.16	300	79	0.3369522
1	algorithm	2.470000	3541	358.85479	129132.45	300	17	1.0000000
4	beta	2.700000	7808	179.44065	79559.21	300	3	0.0000000
7	elitistants	2.883333	2019	195.15145	79343.36	300	14	1.0000000
13	rho	2.936667	7198	138.56366	59061.65	300	4	0.0000000
5	dlb	3.307467	1928	575.36654	242043.01	299	72	1.0000000
3	ants	3.316667	7102	116.72382	51027.76	300	1	0.0000000
11	q0	3.453333	2730	105.42029	44777.97	300	9	1.0000000
6	dummy	3.513333	5974	65.49320	30917.12	300	0	0.0000000
12	rasrank	5.040000	2069	19.67783	12495.60	300	0	1.0000000

Iteration 2:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
9	localsearch	1.910000	1835	459.98740	159663.643	300	73	1
8	instance	2.000000	14049	356.85610	191489.815	300	0	0
1	algorithm	2.416667	2341	282.05808	89777.785	300	32	1
2	alpha	2.416667	6804	260.66572	100297.535	300	4	0
10	nnls	2.586667	5222	295.99198	124105.230	300	76	0
3	ants	2.693333	6663	188.19179	74521.411	300	3	0
7	elitistants	2.704867	1593	167.16038	62511.916	299	24	1
13	rho	3.066667	6253	100.84219	38970.719	300	1	0
4	beta	3.096667	6445	59.84469	29930.291	300	0	0
5	dlb	3.273333	1517	260.68287	97505.161	300	65	1
11	q0	3.370000	2830	170.71023	57873.144	300	18	1
6	dummy	3.380000	5209	57.14056	22998.029	300	4	0
12	rasrank	5.250000	1743	15.92067	7439.316	300	0	1

Iteration 3:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
9	localsearch	1.406667	1313	545.19564	266017.234	300	94	1.0000000
5	dlb	1.730000	979	397.59013	182313.549	300	83	1.0000000
10	nnls	2.090000	6394	338.76438	150586.009	300	40	0.0000000
11	q0	2.156667	5067	333.25243	141953.707	300	31	0.9999786
8	instance	2.286667	17794	206.33843	190000.490	300	0	0.0000000
1	algorithm	2.636667	1987	202.36751	73036.479	300	19	1.0000000
4	beta	2.863333	7253	153.09677	78372.417	300	14	0.0000000
2	alpha	2.920000	7358	112.46777	61591.520	300	0	0.0000000
13	rho	3.120000	6864	101.79529	52342.158	300	5	0.0000000
3	ants	3.213333	6813	107.20464	56919.624	300	1	0.0000000
6	dummy	3.390000	5886	97.99187	39768.468	300	2	0.0000000
7	elitistants	4.014267	654	32.17029	18504.235	276	11	1.0000000
12	rasrank	4.910889	1212	10.44417	8726.568	295	0	1.0000000

Iteration 4:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
9	localsearch	1.806667	1101	291.527826	159854.089	300	60	1.0000000
8	instance	2.063333	19602	133.936001	223541.550	300	0	0.0000000
11	q0	2.213333	6092	344.708999	182533.643	300	83	0.0000000
10	nnls	2.436667	6975	265.924272	140855.976	300	14	0.0000000
1	algorithm	2.460000	912	228.180314	112580.511	292	76	1.0000000
3	ants	2.566667	7152	348.674748	146519.656	300	13	0.0000000
5	dlb	2.600000	1015	211.497903	112547.080	300	31	1.0000000
2	alpha	2.990000	7490	279.046088	137727.225	300	5	0.0000000
4	beta	3.020000	7491	143.553084	84901.974	300	1	0.0000000
13	rho	3.293333	6958	106.601592	59394.786	300	0	0.0000000
6	dummy	3.446667	5621	98.268972	42588.431	300	3	0.0844534
7	elitistants	3.470000	794	46.733525	23985.936	292	14	1.0000000
12	rasrank	5.132500	585	7.443385	6630.966	265	0	1.0000000

Iteration 5:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
2	alpha	1.306667	9689	796.473686	482560.650	300	86	0.0000000
3	ants	1.433333	8780	808.820337	391578.356	300	82	0.0000000
10	nnls	2.176667	8313	531.103247	310765.248	300	10	0.0000000
11	q0	2.236667	8784	381.888064	254350.367	300	54	0.0000000
8	instance	2.473333	27473	67.631619	396629.223	300	0	0.0000000
12	rho	2.773333	8657	328.399675	208774.537	300	21	0.0000000
4	beta	2.826667	9298	319.784186	224320.125	300	0	0.0000000
6	dummy	3.166667	7320	256.041124	153672.199	300	0	0.9888505
1	algorithm	4.135333	501	47.200602	28164.332	260	31	1.0000000
9	localsearch	4.304033	520	39.004017	36213.649	281	8	1.0000000
5	dlb	4.776000	575	37.758168	36473.687	280	8	1.0000000
7	elitistants	6.264067	192	1.321074	1653.337	142	0	1.0000000

Iteration 6:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
7	q0	1.810000	4840	94.24916	37084.24	300	58	1
6	nnls	1.860000	4394	99.25089	31967.32	300	73	1
4	dummy	1.876667	4064	82.21553	27554.09	300	68	1
3	beta	2.260000	5294	86.80981	31574.13	300	37	1
2	ants	2.310000	4552	52.64727	23579.16	300	29	1
1	alpha	2.336667	4989	67.25469	27152.13	300	20	1
5	instance	2.403333	13024	-12.13454	69154.13	300	4	0
8	rho	2.570000	4894	55.58617	24049.89	300	11	1

Iteration 7:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
4	dummy	2.025922	1030	10.213676	1545.2265	293	70	1
6	nnls	2.179256	1005	5.297680	843.8250	293	60	1
5	instance	2.313333	3340	-1.277253	1044.8231	300	0	0
2	ants	2.564544	970	4.671665	662.2813	289	46	1
8	rho	2.661900	1036	3.901766	537.1403	291	46	1
7	q0	2.678511	1027	3.678020	551.6137	286	40	1
1	alpha	2.899200	951	2.402613	394.9327	288	16	1
3	beta	2.909944	1071	2.855522	448.4815	295	22	1

Iteration 8:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
5	instance	1.431507	1696	-0.1556767	175.848108	292	73	0.0000000
7	q0	2.436415	509	0.1650302	14.128630	245	45	0.9999679
4	dummy	2.565069	476	0.1064153	11.993353	232	42	1.0000000
8	rho	2.609817	479	0.0507708	12.077270	234	39	1.0000000
2	ants	2.654338	423	0.0273686	11.129151	234	39	1.0000000
1	alpha	2.742580	474	0.0762815	10.997649	227	20	1.0000000
3	beta	2.807877	439	0.0831447	9.994890	229	33	1.0000000
6	nnls	3.518721	288	0.0771947	5.615506	186	9	1.0000000

The 5 most important parameters per iteration are :

```
## [[1]]
## [1] "alpha"          "localsearch" "instance"      "algorithm"
##
## [[2]]
## [1] "localsearch" "instance"      "algorithm"      "alpha"
##
## [[3]]
## [1] "localsearch" "dlb"          "instance"
##
## [[4]]
## [1] "localsearch" "instance"      "algorithm"      "ants"
##
## [[5]]
## [1] "alpha"        "ants"         "nnls"          "instance"
##
## [[6]]
## [1] "q0"          "nnls" "beta" "ants"
##
## [[7]]
## [1] "instance" "nnls"      "ants"      "rho"
##
## [[8]]
## [1] "instance" "q0"        "rho"        "ants"
```

Parameter relevant interactions:

Iteration 1:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth
alpha	localsearch	1.039007	531	localsearch:alpha	1.573333

Iteration 2:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth
alpha	algorithm	1.136364	527	algorithm:alpha	1.59

Iteration 3:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth
----------	---------------	----------------	-------------	-------------	-----------------------

Iteration 4:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth
----------	---------------	----------------	-------------	-------------	-----------------------

Iteration 5:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth
ants	nnls	0.975945	556	nnls:ants	1.620000
alpha	nnls	1.253700	559	nnls:alpha	1.493333

Iteration 6:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth
----------	---------------	----------------	-------------	-------------	-----------------------

Iteration 7:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth
----------	---------------	----------------	-------------	-------------	-----------------------

Iteration 8:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth
----------	---------------	----------------	-------------	-------------	-----------------------