Supplementary material: A quantile-based approach for hyperparameter transfer learning

David Salinas, Huibin Shen, Valerio Perrone
Amazon
Berlin, Germany
david.salinas.pro@gmail.com, {huibishe, vperrone}@amazon.com

1 Additional figures and tables

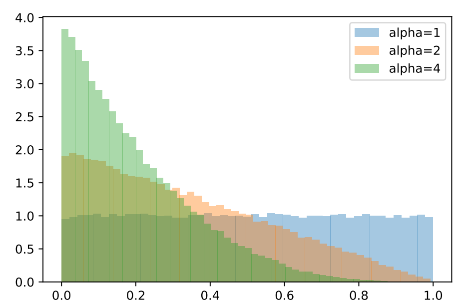

Figure 1: Illustration of the exploration-exploitation trade-off. The plot shows the histogram of 100K quantiles sampled with $p_{\beta(u)}$ for $\beta = 1, 2, 4$. Exploitation increases with β as lower quantiles are likely to be sampled with large values of β. Conversely, $\beta = 1$ leads to pure exploration as all quantiles are sampled uniformly.

tasks	# datasets	# hyperparameters	# evaluations per dataset	metrics available
DeepAR	11	6	~ 220	quantile loss, time
FCNET	4	9	62208	MSE, time
XGBoost	10	9	5000	1-AUC

Table 1: A summary of the three HPO problems we considered.

3rd Workshop on Meta-Learning at NeurIPS 2019, Vancouver, Canada.

tasks	hyperparameter	search space	type	scale
DeepAR	# layers	$[1, 5]$	integer	linear
	# cells	$[10, 120]$	integer	linear
	learning rate	$[10^{-4}, 0.1]$	continuous	log10
	dropout rate	$[10^{-2}, 0.5]$	continuous	log10
	context_length_ratio	$[10^{-1}, 4]$	continuous	log10
	# bathes per epoch	$[10, 10^{4}]$	integer	log10
XGBoost	num_round	$[2, 2^{9}]$	integer	log2
	eta	$[0, 1]$	continuous	linear
	gamma	$[2^{-20}, 2^{6}]$	continuous	log2
	min_child_weight	$[2^{-8}, 2^{6}]$	continuous	log2
	max_depth	$[2, 2^{7}]$	integer	log2
	subsample	$[0.5, 1]$	continuous	linear
	colsample_bytree	$[0.3, 1]$	continuous	linear
	lambda	$[2^{-10}, 2^{8}]$	continuous	log2
	alpha	$[2^{-20}, 2^{8}]$	continuous	log2
FCNET	initial_lr	$\{0.005, 0.001, 0.05, 0.01, 0.05, 0.1\}$	categorical	-
	batch_size	$\{8, 16, 32, 64\}$	categorical	-
	lr_schedule	{cosine, fix}	categorical	-
	activation layer 1	{relu, tanh}	categorical	-
	activation layer 2	{relu, tanh}	categorical	-
	size layer 1	$\{16, 32, 64, 128, 256, 512\}$	categorical	-
	size layer 2	$\{16, 32, 64, 128, 256, 512\}$	categorical	-
	dropout layer 1	$\{0.0, 0.3, 0.6\}$	categorical	-
	dropout layer 2	$\{0.0, 0.3, 0.6\}$	categorical	-

Table 2: A summary of the search spaces for the three algorithms.

task	dataset	ABLR	GP	RS	quantile-GP	quantile-RS
DeepAR	electricity	-7.4	-4.0	0.0	-0.7	-3.8
	exchange-rate	-0.4	4.6	0.0	7.4	4.0
	m4-Daily	-3.4	-6.2	0.0	1.8	-9.2
	m4-Hourly	-35.7	-5.5	0.0	1.8	-21.7
	m4-Monthly	-2.1	1.9	0.0	0.9	0.2
	m4-Quarterly	-1.2	1.6	0.0	-6.7	-8.4
	m4-Weekly	-12.5	0.3	0.0	-8.3	-0.7
	m4-Yearly	-0.5	-7.7	0.0	0.8	0.8
	solar	-1.7	1.4	0.0	1.3	0.4
	traffic	-4.0	-0.8	0.0	-0.4	-3.7
	wiki-rolling	0.4	0.8	0.0	1.0	-0.2
FCNet	naval	64.3	-12.1	0.0	74.3	66.2
	parkinsons	16.7	21.7	0.0	41.6	29.9
	protein	-1.8	7.4	0.0	11.3	4.2
	slice	-0.3	28.1	0.0	51.1	44.3
XGBoost	a6a	-0.6	0.0	0.0	0.1	0.0
	australian	-3.7	0.7	0.0	4.4	1.0
	german.numer	-2.4	-0.8	0.0	-0.3	0.3
	heart	1.0	4.5	0.0	3.2	-1.8
	ijcnn1	-20.4	2.5	0.0	3.1	-1.6
	madelon	2.2	9.0	0.0	7.3	-1.7
	spambase	-7.5	1.9	0.0	2.4	1.4
	svmguide1	-6.5	0.2	0.0	-0.5	-1.7
	w6a	0.0	4.4	0.0	5.1	1.8

Table 3: Relative improvements over random search averaged over iterations.

task	dataset	ABLR	GP	RS	quantile-GP	quantile-RS
DeepAR	electricity	-12.3	-5.3	0.0	2.9	5.0
	exchange-rate	6.6	8.3	0.0	13.2	28.7
	m4-Daily	-9.8	-6.4	0.0	7.2	10.0
	m4-Hourly	-66.6	2.2	0.0	10.9	19.8
	m4-Monthly	2.7	5.0	0.0	5.7	4.9
	m4-Quarterly	-7.0	0.2	0.0	-0.2	-0.7
	m4-Weekly	-19.4	-2.1	0.0	-3.8	-0.8
	m4-Yearly	-2.2	-14.3	0.0	1.0	1.6
	solar	-0.3	1.9	0.0	2.2	3.3
	traffic	-26.7	1.3	0.0	3.7	2.2
	wiki-rolling	-0.9	0.0	0.0	1.3	2.6
FCNet	naval	2.2	-78.3	0.0	-1.3	64.9
	parkinsons	-14.3	19.6	0.0	30.6	27.2
	protein	-4.4	7.3	0.0	8.0	5.7
	slice	-12.0	28.0	0.0	41.9	53.6

Table 4: Relative improvements over random search averaged over time.