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

David Salinas
NAVER LABS Europe [*]
david.salinas@naverlabs.com

Huibin Shen
Amazon Research
huibishe@amazon.com

Valerio Perrone
Amazon Research
vperrone@amazon.com

1 Additional figures and tables

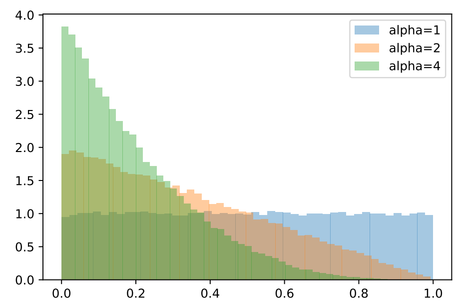

Figure 1: Illustration of the exploration-exploitation trade-off. The plot shows the histogram of 100K quantiles sampled with $p_{\beta(u)}$ for $\beta = 1, 2, 4$. Exploitation increases with β as lower quantiles are likely to be sampled with large values of β. Conversely, $\beta = 1$ leads to pure exploration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

[*]Work done while being at Amazon Research

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.

blackbox	dataset	ABLR	GP	Quantile-GP	Quantile-RS	RS
DeepAR	electricity	-0.76	0.19	**0.40**	-0.13	0.00
	exchange-rate	0.50	0.17	**0.64**	0.15	0.00
	m4-Daily	-0.51	-0.16	**0.32**	-0.09	0.00
	m4-Hourly	-1.11	-0.24	-0.02	-0.29	**0.00**
	m4-Monthly	-0.73	0.35	**0.78**	-0.18	0.00
	m4-Quarterly	-0.52	0.17	**0.55**	-0.31	0.00
	m4-Weekly	-0.36	0.22	**0.61**	0.19	0.00
	m4-Yearly	0.04	0.70	**0.75**	0.23	0.00
	solar	-0.06	0.21	**0.36**	0.02	0.00
	traffic	-0.69	0.31	**0.46**	0.01	0.00
	wiki-rolling	-0.14	0.01	**0.07**	-0.21	0.00
fcnet	naval	0.69	-0.59	**0.99**	0.86	0.00
	parkinsons	0.32	0.47	**0.87**	0.54	0.00
	protein	0.07	0.41	**0.92**	0.59	0.00
	slice	0.03	0.42	**0.97**	0.75	0.00
xgboost	a6a	-0.23	0.42	**0.44**	0.00	0.00
	australian	-0.11	0.20	**0.34**	-0.02	0.00
	german.numer	-0.26	-0.05	**0.12**	**0.12**	0.00
	heart	0.13	0.20	**0.26**	-0.02	0.00
	ijcnn1	-0.83	0.52	**0.69**	0.21	0.00
	madelon	0.37	0.63	**0.65**	-0.08	0.00
	spambase	-0.42	0.34	**0.41**	0.09	0.00
	svmguide1	-0.81	**0.31**	0.27	0.14	0.00
	w6a	0.15	0.61	**0.66**	0.15	0.00

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

blackbox	dataset	ABLR	GP	Quantile-GP (time)	Quantile-RS (time)	RS
DeepAR	electricity	-1.54	0.01	**0.53**	0.28	0.00
	exchange-rate	-0.11	-0.02	**0.88**	0.87	0.00
	m4-Daily	-0.36	-0.08	-0.08	**0.71**	0.00
	m4-Hourly	-0.14	0.00	**0.71**	0.61	0.00
	m4-Monthly	0.64	-0.00	**0.79**	0.72	0.00
	m4-Quarterly	-2.87	-0.08	**0.19**	-0.09	0.00
	m4-Weekly	-3.17	-0.01	**0.26**	0.24	0.00
	m4-Yearly	-0.57	-0.09	**0.31**	0.28	0.00
	solar	0.42	-0.01	**0.57**	0.54	0.00
	traffic	-2.39	-0.01	**0.73**	0.69	0.00
	wiki-rolling	-0.64	0.01	**0.56**	0.55	0.00
fcnet	naval	-1.84	-2.27	**0.94**	0.89	0.00
	parkinsons	-0.63	0.27	**0.73**	0.62	0.00
	protein	-0.08	0.00	**0.42**	0.36	0.00
	slice	0.20	0.00	**0.93**	0.74	0.00

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

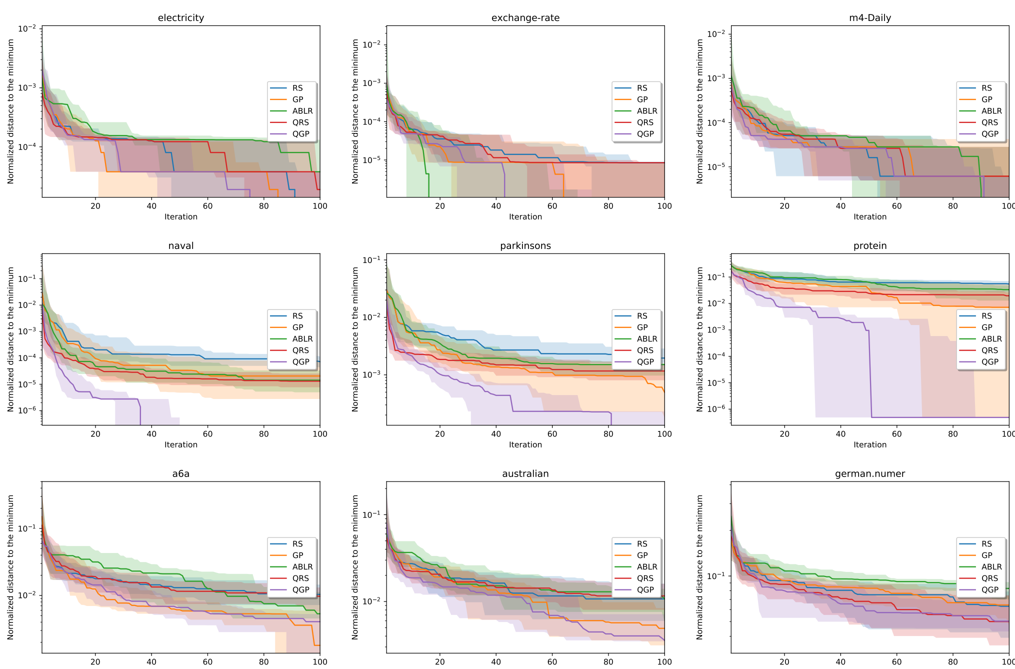

Figure 2: Error over iteration for the 3 first datasets of each blackbox for all methods. Confidence intervals are computed with P20, P50 and P80 quantiles on 30 random seeds.