A quantile-based approach for hyperparameter transfer learning

David Salinas[2] Huibin Shen[1] Valerio Perrone[1]

[1]Amazon Research
[2]NAVER LABS Europe, work done at Amazon

December 11, 2019

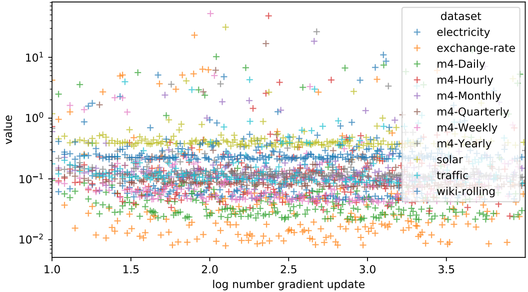
Transfer learning setting

- Assume many HP evaluations $\{x_i^l, y_i^l\}_{i=0}^{n_l}$ available for n_l datasets
- $x_i^l \in \mathbb{R}^d$ hyperparameter, $y_i^l \in \mathbb{R}$ objective to be minimized
- Can we use it to speed up the tuning of a new dataset?

Transfer learning

Difficulties:

- Scales of objectives y_i^l may vary significantly across tasks
- Noise may not be Gaussian
- Many observations: hard to apply (approximate) GP



Gaussian Copula transform

If only every y^l was Gaussian...

- Apply change of variable $\psi = \Phi^{-1} \circ \mathsf{F}$
- Φ Gaussian CDF, F is the marginal CDFs (approximated with empirical CDF)
- $z^l = \psi(y^l)$
- All z^l becomes centered Gaussian! $z^l \in \mathcal{N}(0, 1)$

Transfer learning

Parametric Prior

- Regress $z(x) \approx \mathcal{N}(\mu_\theta(x), \sigma_\theta(x))$
- Parameters θ are learned with MLE on evaluations
- *Joint-learning* as θ *tied across tasks (only possible because* z *have comparable scales across tasks !)*

Two HPO strategies

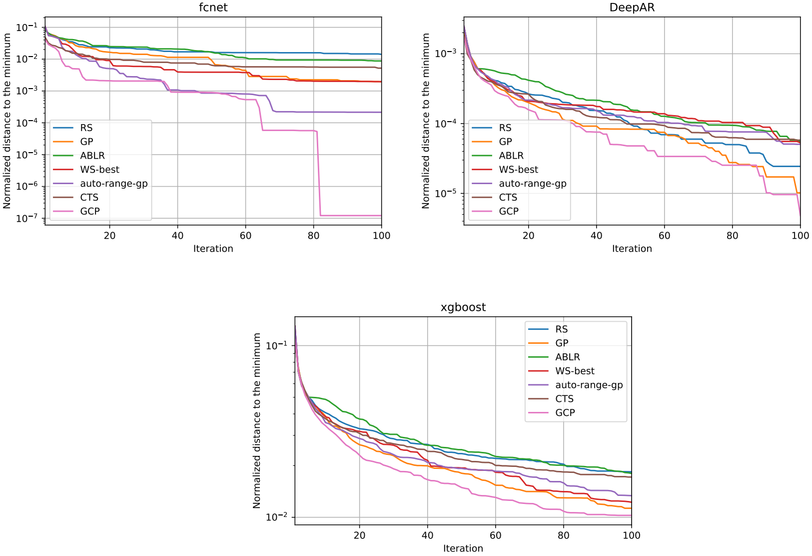
- Thompson sampling with $\mathcal{N}(\mu_\theta(x), \sigma_\theta(x))$
- Gaussian Copula Process with the prior $\mathcal{N}(\mu_\theta(x), \sigma_\theta(x))$

Results

- Evaluate on 3 blackboxes with precomputed evaluations (MLP [Klein 18], DeepAR [Salinas 17], XGboost)

blackbox	# datasets	# hyperparameters	# evaluations	objectives
DeepAR	11	6	$\sim$ 220	quantile loss, time
FCNET	4	9	62208	MSE, time
XGBoost	9	9	5000	1-AUC

Results


fcnet
Normalized distance to the minimum
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
20
40
60
80
100
Iteration
RS
GP
ABLR
WS-best
auto-range-gp
CTS
GCP
DeepAR
Normalized distance to the minimum
10^{-3}
10^{-4}
10^{-5}
20
40
60
80
100
Iteration
RS
GP
ABLR
WS-best
auto-range-gp
CTS
GCP
xgboost
Normalized distance to the minimum
10^{-1}
10^{-2}
20
40
60
80
100
Iteration
RS
GP
ABLR
WS-best
auto-range-gp
CTS
GCP

Results

- Because every objectives are Gaussian centered, we can easily combined them!
- Multi-objective: optimize accuracy/time trade-off with $z^{\text{error}}(x) + z^{\text{runtime}}(x)$
- More at our poster!