A Channel Coding Benchmark for Meta-Learning

Samsung Research

Email: rui.li@samsung.com
Website: https://ruihuili.github.io

WeChat

Rui Li, Ondrej Bohdal, Hyeji Kim, Da Li, Nicholas D. Lane, and Timothy Hospedales

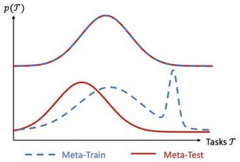
Challenges and Contributions

Meta-learners have been shown to suffer in realistic settings [1,2], especially when:

- Task distribution is broad and multi-modal
- There is distribution shift between the meta-training and meta-testing tasks.

Our Contributions:

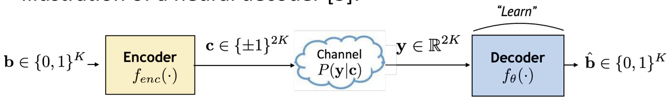
- Investigate the effects of the **diversity** of task distributions, and **shift** between meta-train and meta-test on performance.
- Introduce quantitative metrics of task-distribution shift and training-data diversity score.



Channel Coding

- Fundamental problem in communications;
- Practical application where task distributions naturally arise, and fast adaptation to new tasks is practically valuable.

Illustration of a neural decoder [3]:



References:

[1] Triantafillou, Eleni, et al. "Meta-dataset: A dataset of datasets for learning to learn from few examples." arXiv preprint arXiv:1903.03096 (2019).

[2] Yu, Tianhe, et al. "Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning." Conference on Robot Learning. PMLR, 2020.

[3] Kim, Hyeji, et al. "Communication algorithms via deep learning." NeurIPS, 2018

[4] Arnold, Sebastien M. R., et al. "learn2learn: A Library for Meta-Learning Research."

Coding Benchmark for Meta-Learning

4 Families (modes) of **channel models**: AWGN, Bursty, Memory, and Multipath interference channels, and corresponding decoding tasks.

A **task distribution** corresponds to a channel class and is **parameterized** by continuous channel parameters ω, e.g., SNR value.

Implementation: Based on and extended Learn2Learn [4] framework.

Definition 1: Train-Test Task-Shift $S(p_a(\mathcal{T}), p_b(\mathcal{T}))$
Distance between a test distribution $\mathcal{T}_a$ and a train distribution $\mathcal{T}_b$ *using KLD:*

$$S(p_a(\mathcal{T}), p_b(\mathcal{T})) := \mathbb{E}_{\mathbf{c}}[D_{KL}(p_a(\mathbf{y}_a|\mathbf{c}) || p_b(\mathbf{y}_b|\mathbf{c}))] + \mathbb{E}_{\mathbf{c}}[D_{KL}(p_b(\mathbf{y}_b|\mathbf{c}) || p_a(\mathbf{y}_a|\mathbf{c}))],$$

where $p_a(y_a|c)$ and $p_b(y_b|c)$ denote the channels associated with $\mathcal{T}_a$ and $\mathcal{T}_b$, respectively.

Definition 2: The Diversity Score $D(\mathcal{T})$ of a task distribution $p(\mathcal{T})$ is defined as mutual information between the channel parameter ω and the received signal y:

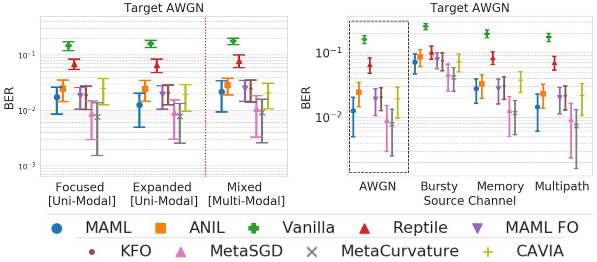
$$D(\mathcal{T}) = \mathbb{E}_{\mathbf{c}}[I(\omega; \mathbf{y}|\mathbf{c})],$$

Where ω denotes the channel parameter (latent variable) for the task distribution, i.e.

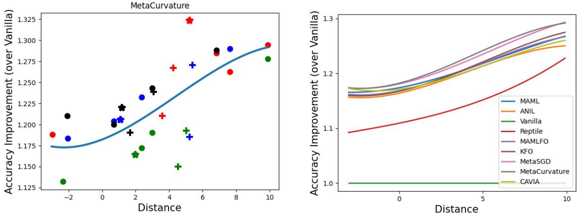
$$p(\mathbf{y}|\mathbf{c}) = \int_\omega p(\mathbf{y}|\mathbf{c}, \omega) p_\omega(\omega).$$

Experiments

Impact of Training Distribution **Diversity** (Left) and Train-Test **Distribution Shift** (Right) on Meta-Learning Performance:



Proposed **Distance Score** vs **Accuracy** Improvement over Vanilla (Non-Meta-ERM):



Take Home Messages:

- Mild degradation in performance under complex task distributions
- Absolute performance degrades rapidly with distribution shift.
- Accuracy improvement over non-meta-learner improves with shift
- Channel coding provides a flexible benchmark for studying meta-learning

*Who is taking the **feature re-use** short-cut?* 🔍