Provable General Function Class Representation Learning in Multitask Bandits and MDP

Rui Lu[1], Andrew Zhao[1], Simon Du[2], Gao Huang[1]

[1]Tsinghua University
[2]University of Washington

Recent Progress in Multitask Learning

- Multitask representation learning (MRL) is widely used in practice to reduce sample complexity.

- Theoretical understanding for its mechanism is still limited.
- Most previous works [Yang et al. 2021, Hu et al. 2021, Lu et al. 2021] can only deal with linear or known representation, which is far from real scenarios.

Why and how MRL can effectively reduce sample complexity?

Can we analyze it in complex general function class?

Our Result

NEURAL INFORMATION PROCESSING SYSTEMS

- We propose a simple and straightforward algorithm called GFUCB (Generalized Functional Upper Confidence Bound) for multitask contextual bandits and MDPs.

Main theorem for GFUCB in multitask bandits

GFUCB achieves regret bound as $\tilde{O}\left(\sqrt{MdT(Mk + \log N(\Phi, \alpha_T, \|\cdot\|_\infty))}\right)$

M: number of tasks, d: eluder dimension of value function class,
k: dimension of representation, T: total number of steps,
$N(\Phi, \alpha_T, \|\cdot\|_\infty)$: covering number of function class Φ.

- Extract unknown general representation function instead of being given one [Jin et al 2020].
- Same optimal as [Hu et al. 2021] when Φ is linear class.

Our Result (cont'd)

- Similar results holds for low inherent Bellman Error MDP.

Main theorem for GFUCB in multitask linear MDP
For multitask linear MDP, GFUCB achieves regret bound as $\tilde{O}\left(MH\sqrt{Tdk} + H\sqrt{MdT\log N(\Phi, \alpha_T, \|\cdot\|_\infty)} + MHTI\sqrt{d}\right)$
H: planning horizon, I: the inherent Bellman error

- The dominating term (red) is sublinear in number of tasks M,
- More tasks, lower average regret.

Mechanism Explained

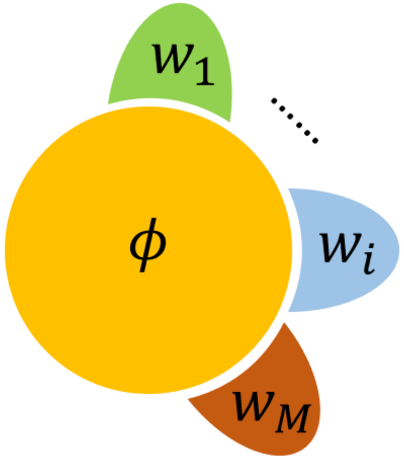
$$\mathcal{F}^M = (\mathcal{L} \circ \Phi)^M$$

- Complex, many parameters to learn.
- Tasks are independent, inefficient.

Mechanism Explained

$$\mathcal{F}^M = (\mathcal{L} \circ \Phi)^M$$

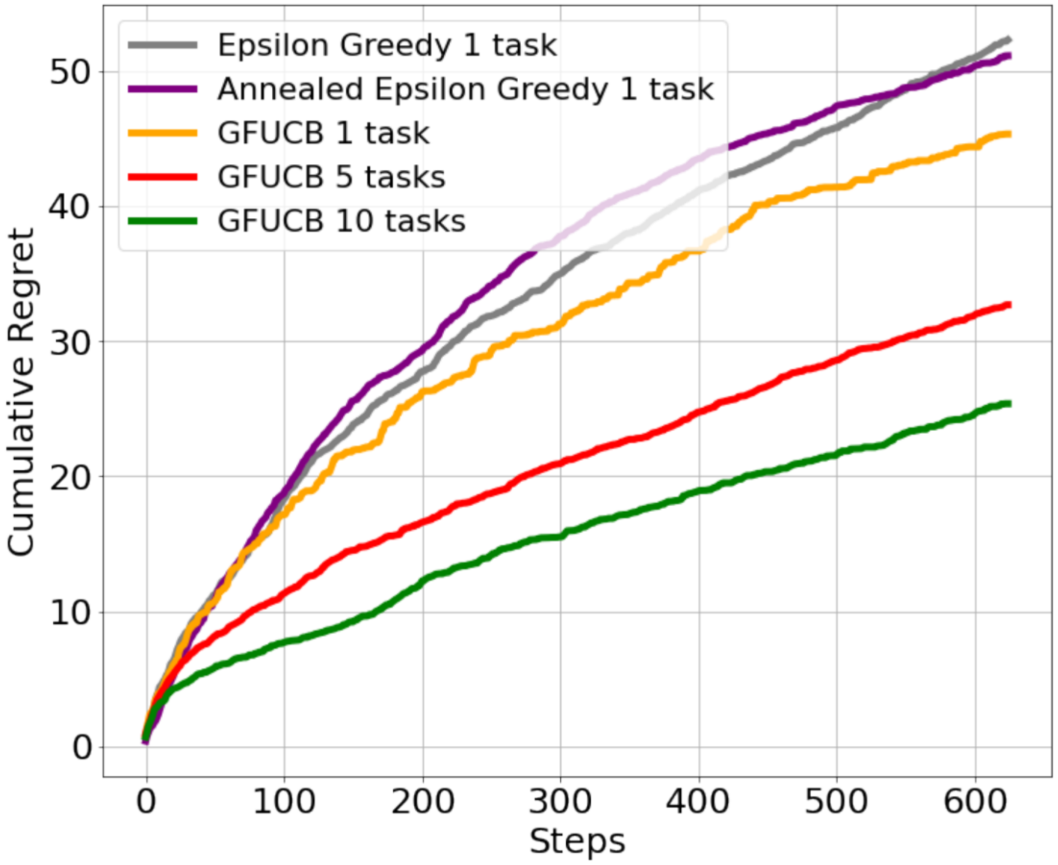
- Complex, many parameters to learn.
- Tasks are independent, inefficient.



$$\mathcal{F}^{\otimes M} = \mathcal{L}^M \circ \Phi$$

- More compact, much less parameters.
- Star structure. All tasks share the same representation backbone ϕ.
- Require much fewer sample to train.

Experimental Result



- GFUCB is better than naïve exploration.
- The sample efficiency is scalable to the number of tasks.

Conclusion

NEURAL INFORMATION PROCESSING SYSTEMS

- General function multitask representation learning is also feasible and efficient.
- Multitask training uses samples from all the tasks to jointly find shared knowledge as a representation ϕ, which accelerates each task's convergence.
- Technical contribution: multihead function class for analyzing multitask representation learning.

- For more details: Please refer to our paper

https://arxiv.org/pdf/2205.15701

NEURAL INFORMATION PROCESSING SYSTEMS

Thanks for Listening