Deep Online Learning via Meta-Learning: Continual Adaptation for Model-Based RL

Anusha Nagabandi, Chelsea Finn, Sergey Levine
UC Berkeley

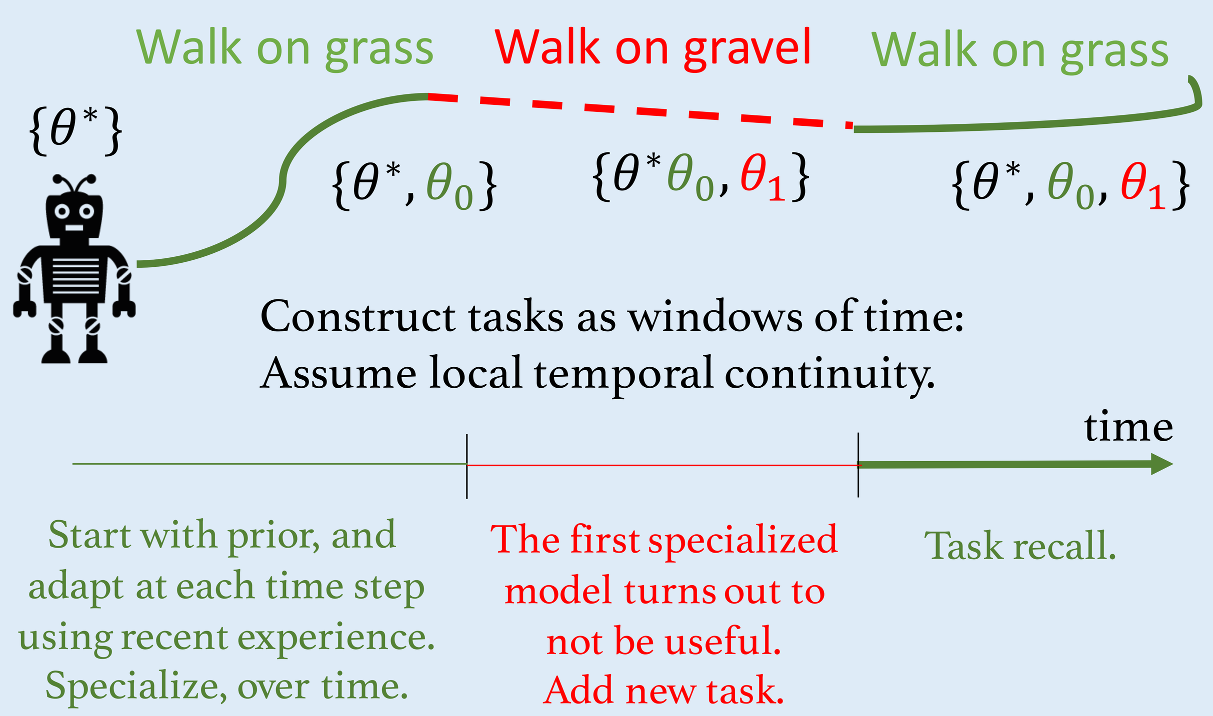
Motivation

Neural networks can represent complex functions, but lack the ability to **learn online** to handle the non-stationarity of the real world.

SGD for direct **online adaptation** of large function approximators is not effective.

Animals/humans remember experiences and **recall** them to adapt more quickly to similar disturbances in the future.

Online Learning: Problem Setup

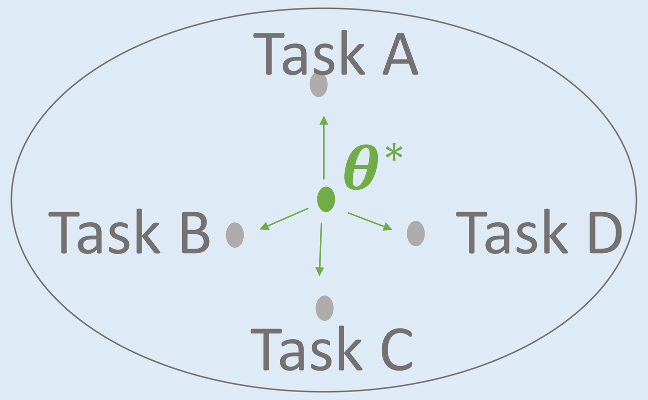


Meta-Learning a Model that can Adapt

The model-agnostic meta-learning (MAML, Finn et al. 2017) training objective over data from tasks T is defined as :

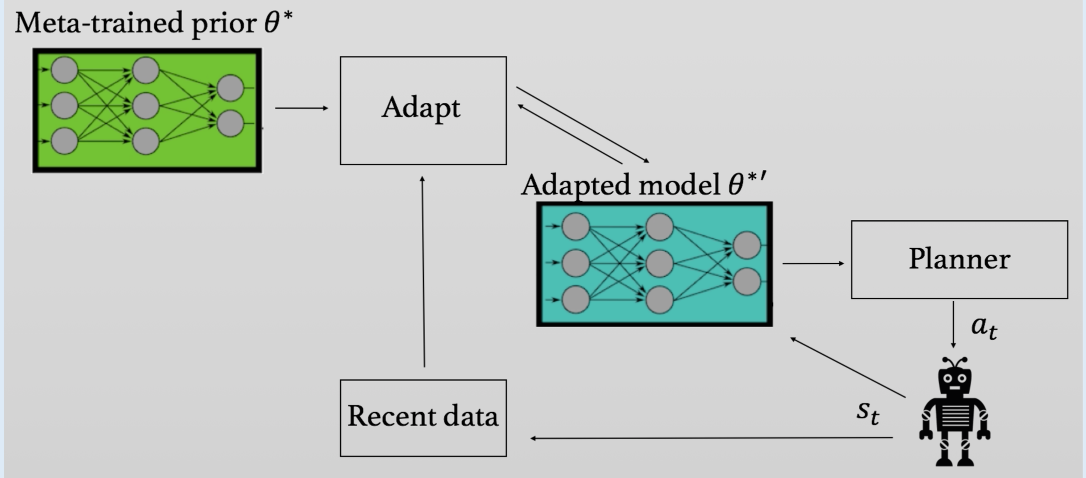
$$\theta^* = \underset{\theta}{\text{argmin}} \sum_{T} L_{\theta'_T}(D_T^{val})$$

where $\theta'_T = \theta - \alpha \nabla_\theta L_\theta(D_T^{tr})$



Resulting θ^* acts as a prior from which we fine-tune at run-time:

$$\theta^* - \alpha \nabla_{\theta^*} L_{\theta^*}(D_{T_{test}}^{tr})$$



Prior work (Nagabandi et al. 2018) resets to prior at every time step.

Online Learning with Mixture of Networks

Problem:
Infer latent task variable T_t at each time step t, while continually adapting mixture of models $\theta(T_i)\ \forall i$

Initialize task distribution θ

$|T| = 1$
$\theta = \{\theta^*\}$

Calculate **likelihood** of recently seen data under each of the tasks

$$p_{\theta(T_i)}(\boldsymbol{y_t} | \boldsymbol{x_t}, T_t = T_i)$$

Check if
$p(data\,|T_{new}) > p(data\,|T_i)$

Add a new task to θ

Get $\boldsymbol{x_{t+1}}$

Expectation:
Estimate latent task probabilities

$$P(T_t = T_i | \boldsymbol{x_t}, \boldsymbol{y_t}) = \underbrace{p_{\theta(T_i)}(\boldsymbol{y_t}|\boldsymbol{x_t}, T_t = T_i)}_{\text{Likelihood}}\ \underbrace{P(T_t = T_i)}_{\text{Prior (CRP)}}$$

Maximization:
Maximize log likelihood of data, with update step size proportional to corresponding task probability

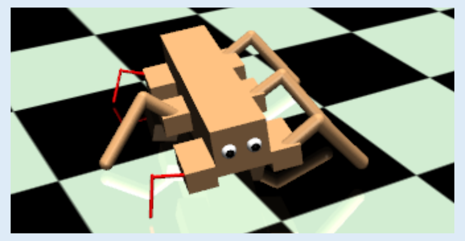
$$\text{Update } \theta(T_i) \propto P(T_t = T_i | \boldsymbol{x_t}, \boldsymbol{y_t})$$

Use model with highest corresponding posterior task probability

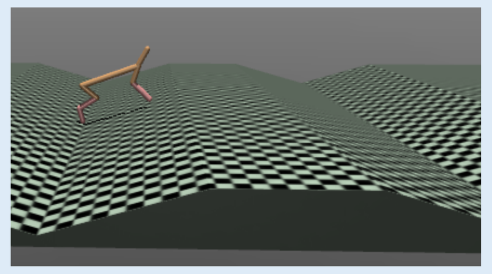
Predict/plan under $p_{\theta(T_{max})}$

Online Learning Experiments for Model-Based RL

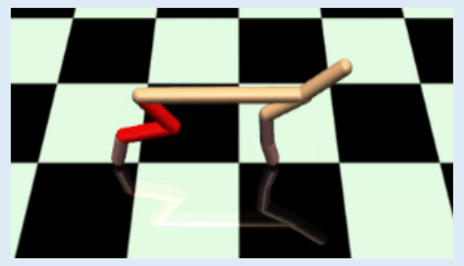
End-effector Crippling



Terrain Slopes



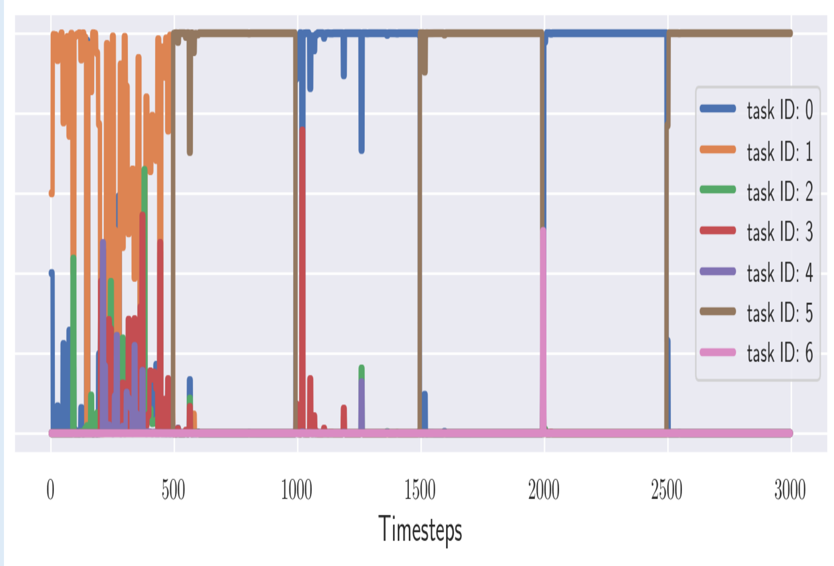
Motor Malfunctions



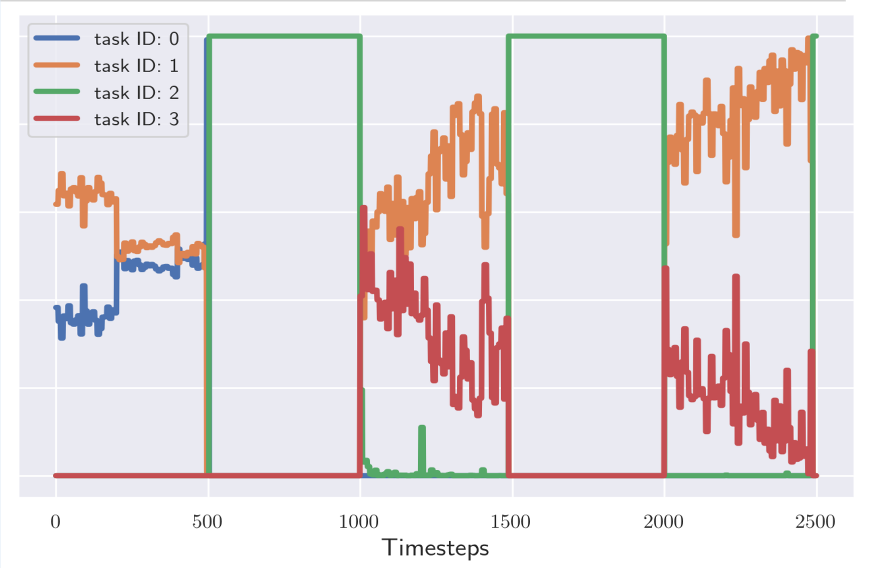
We use our online learning algorithm in a model-based setup to adapt learned dynamics models to system/environment changes, and plan with these constantly updating models

- Input $\boldsymbol{x_t}$: $[s_{t-1}, a_{t-1}]$
- Output $\boldsymbol{y_t}$: s_t

Latent Task Probabilities



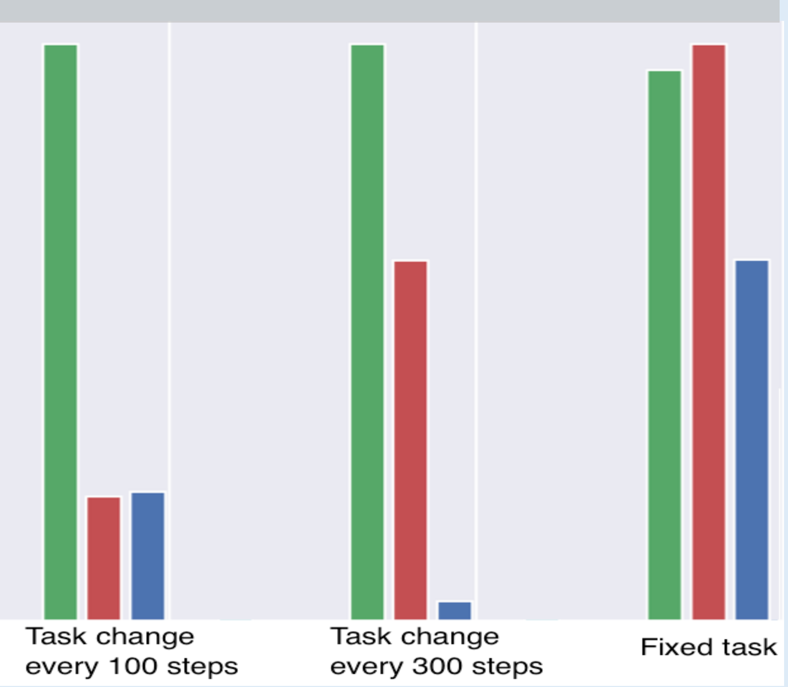
Latent Task Probabilities



- Online task instantiation
- Task specialization, as well as task recall

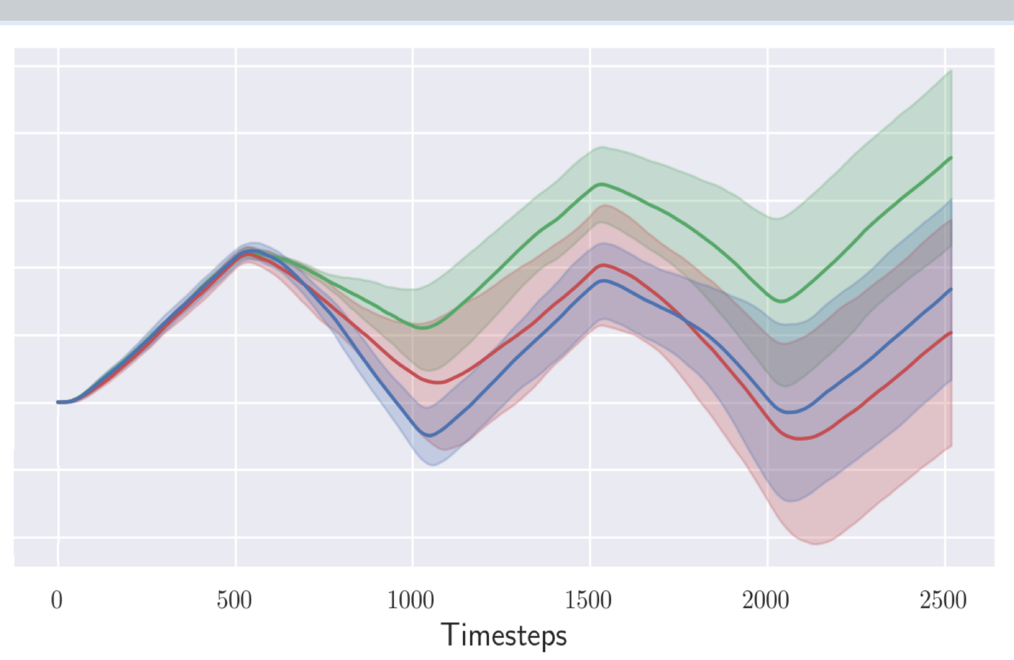
Mixture of Models (ours)
Continuous adaptation
K-shot adaptation from meta-learned model
Standard supervised learning

Performance



- Need adaptation
- K-shot can be insufficient
- Continuous adaptation can be detrimental

Cumulative Sum of Rewards



Takeaways

- Meta-learning enables online-learning with neural networks
- Our EM algorithm can build mixture of models to enable both specialization and recall