PILCO

Xingdong Zuo*

1 Key idea

- Reduce sample complexity: prior knowledge (e.g. demonstration) / extracting more information (e.g. learn dynamics)
- Model bias in model-based RL: use probabilistic model instead of deterministic
- Probabilistic model: GP (sample efficient in low-dim, unscalable) , BNN, ensembles
- PILCO: propagate state distribution analytically via GP and incorporate the uncertainty into planning and policy evaluation.

2 Method

Algorithm 1: PILCO [1]

Randomly initialize the policy π_θ
for *iteration = 1, 2, . . .* **do**
  Execute the system with π_θ and augment the dataset.
  Re-train dynamics model.
  for *optimization iteration = 1, . . . , 1000* **do**
    Predict system trajectories from $p(X_0)$ to $p(X_T)$.
    Evaluate the policy: $J(\theta) = \sum_{t=0}^{T} \gamma^t \mathbb{E}_X \left[\text{cost}(X_t)\right]$.
    Optimize the policy: $\theta \leftarrow \text{argmin}_\theta J(\theta)$
  end
end

Algorithm 2: Deep PILCO [2]

Randomly initialize the policy π_θ
for *iteration = 1, 2, . . .* **do**
  Execute the system with π_θ and augment the dataset.
  Re-train BNN dynamics model.
  for *optimization iteration = 1, . . . , 1000* **do**
    Predict system trajectories from $p(X_0)$ to $p(X_T)$.
    Sample K particles from initial distribution $x_0^k \sim p(X_0)$
    Sample K set of weights for BNN dynamics model $\{W^{(k)}\}_{k=1}^K$
    for *t = 1, . . . , T* **do**
      Obtain output particles $\{y_t^{(k)}\}$ by evaluating $\{W^{(k)}\}$ and $\{x_t^{(k)}\}$ for all $k = 1, \ldots, K$
      Calculate the mean μ_t and standard deviation σ^t of $\{y_t^{(1)}, \ldots, y_t^{(K)}\}$
      Sample new set of K particles $x_{t+1}^{(k)} \sim \mathcal{N}(\mu_t, \sigma_t^2)$
    end
    Evaluate the policy: $J(\theta) = \sum_{t=0}^{T} \gamma^t \mathbb{E}_X \left[\text{cost}(X_t)\right]$.
    Optimize the policy: $\theta \leftarrow \text{argmin}_\theta J(\theta)$
  end
end

*December 18, 2018

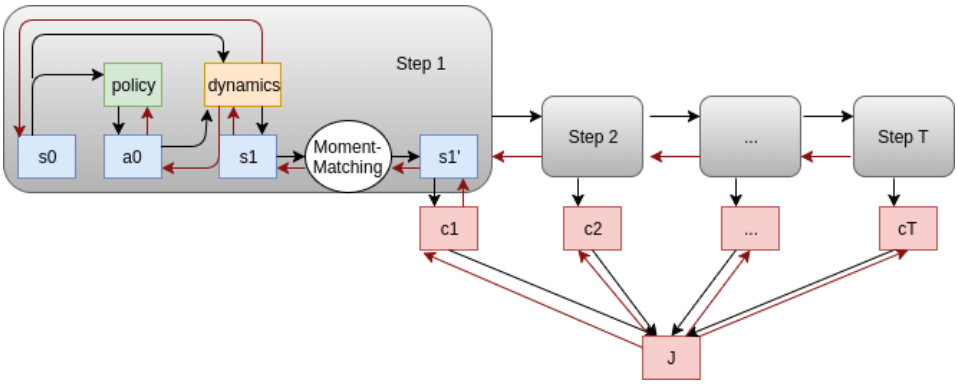

Figure 1: Computational graph of DeepPILCO.

References

[1] Marc Deisenroth and Carl E Rasmussen. "PILCO: A model-based and data-efficient approach to policy search". In: *Proceedings of the 28th International Conference on machine learning (ICML-11)*. 2011, pp. 465–472.

[2] Yarin Gal, Rowan McAllister, and Carl Edward Rasmussen. "Improving PILCO with Bayesian neural network dynamics models". In: *Data-Efficient Machine Learning workshop, ICML*. 2016.