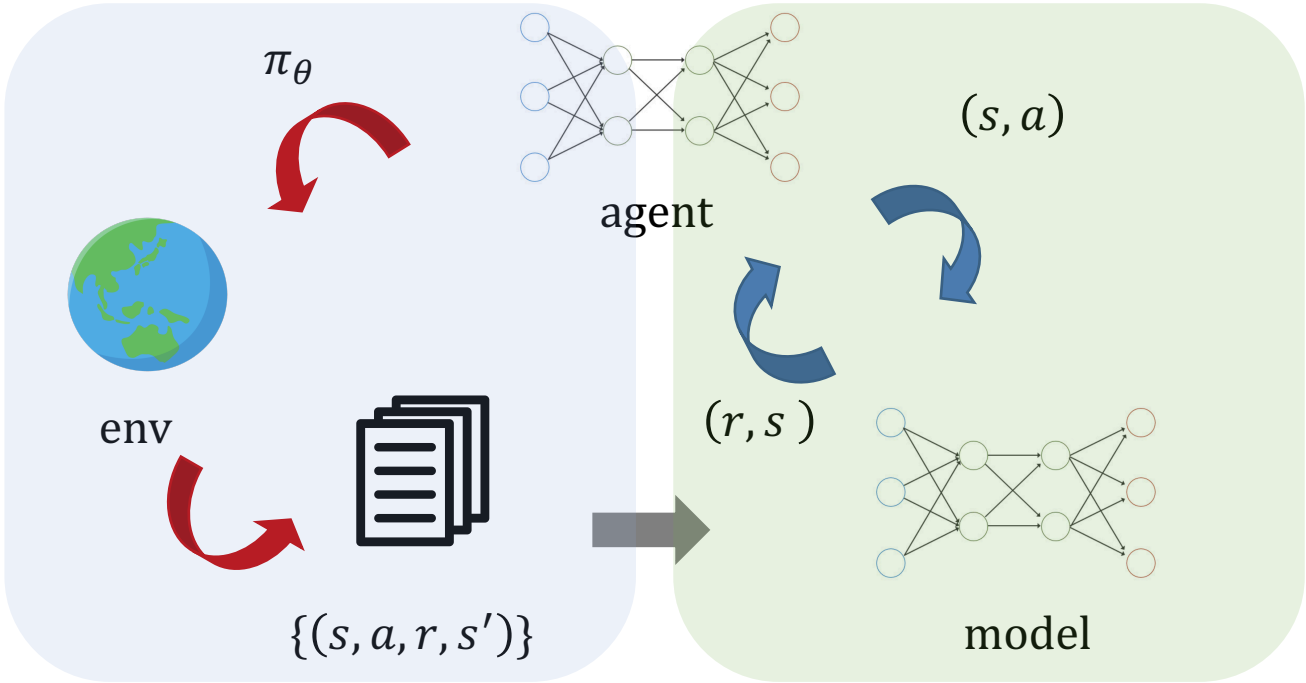
Weekly Report

M2 Ibuki Takeuchi

- Read a paper
  - [1]: T. Matsushima, H. Furuta, Y. Matsuo, O. Nachum and S. S. Gu. “Deployment-Efficient Reinforcement Learning via Model-Based Offline Optimization”, *arXiv preprint arXiv: 2006.03647v2,* 2020.

- [1] is a hybrid RL of offline and online



K times
Data sampling
(K = 1 for
offline RL)

T times
Policy update
(T = 1 for
online RL)

Weekly Report

M2 Ibuki Takeuchi

- Model estimation is suffered from distributional shift
  - How is unknown state? How is unknown input?
  - Policy update enhances input distributional shift

  - Trade-off between performance and policy update (parameter T) ?

- [1] proved it:

$$\eta(\pi) \geq \hat{\eta}(\pi) - C(\varepsilon_m, \varepsilon_\pi)$$

$$C(\varepsilon_m, \varepsilon_\pi) = O(\varepsilon_\pi), \varepsilon_\pi = D_{TV}(\pi_{sample}, \pi)$$

$\varepsilon_\pi \leq O(T)$, so ε_π can be linear for T

It gets harder to guarantee performance as T gets larger

Weekly Report

M2 Ibuki Takeuchi

- Read papers to see approaches with control theory
  - N. M. Yazdani, R. K. Moghaddam, B. Kiumarsi and H. Modares, "A Safety-Certified Policy Iteration Algorithm for Control of Constrained Nonlinear Systems," in *IEEE Control Systems Letters*, vol. 4, no. 3, pp. 686-691, 2020.
  - M. Hertneck, J. Köhler, S. Trimpe and F. Allgöwer, "Learning an Approximate Model Predictive Controller With Guarantees," in *IEEE Control Systems Letters*, vol. 2, no. 3, pp. 543-548, 2018.