Tight Performance Guarantees of Imitator Policies with Continuous Actions

Davide Maran, Alberto Maria Metelli, Marcello Restelli

POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Behavioral Cloning and noise injection

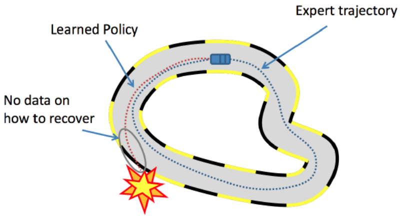
Behavioral cloning

Train policy π_I to minimize $\mathbb{E}_{s \sim d^{\pi_E}}[D(\pi_I(\cdot|s), \pi_E(\cdot|s))]$.

Behavioral Cloning and noise injection

Behavioral cloning

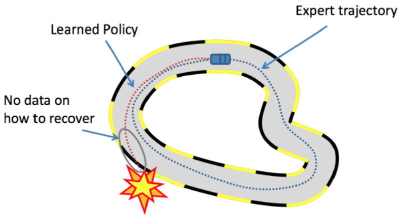
Train policy π_I to minimize $\mathbb{E}_{s \sim d^{\pi_E}}[D(\pi_I(\cdot|s), \pi_E(\cdot|s))]$.



Behavioral Cloning and noise injection

Behavioral cloning

Train policy π_I to minimize $\mathbb{E}_{s \sim d^{\pi_E}}[D(\pi_I(\cdot|s), \pi_E(\cdot|s))]$.



Noise injection

Apply noise on π_E to explore neighbourhood of the trajectory

Empirically validated [Laskey et al. 2017].

Current theoretical results and open problems

Theorem [Xu, Li, and Yu 2020]

π_E = expert policy, π_I = imitator policy.

$$J^{\pi_E} - J^{\pi_I} \leq \frac{2R_{\max}}{(1-\gamma)^2}\mathbb{E}_{s\sim d^{\pi_E}}[TV(\pi_E(\cdot|s), \pi_I(\cdot|s))].$$

Not suited for continuous actions

If π_E deterministic, $TV(\pi_E(\cdot|s), \pi_I(\cdot|s)) = 1$.

Current theoretical results and open problems

Theorem [Xu, Li, and Yu 2020]

π_E = expert policy, π_I = imitator policy.

$$J^{\pi_E} - J^{\pi_I} \leq \frac{2R_{\max}}{(1-\gamma)^2}\mathbb{E}_{s\sim d^{\pi_E}}[TV(\pi_E(\cdot|s), \pi_I(\cdot|s))].$$

Not suited for continuous actions

If π_E deterministic, $TV(\pi_E(\cdot|s), \pi_I(\cdot|s)) = 1$.

Can we explain the performance of Noise injection?

The bound shows no advantage in using noise injection.

Mathematical background

Lipschitz continuity

$f : X \to Y$ is *L-Lipschitz continuous*:

$$d_Y(f(x), f(x')) \leq L d_X(x, x'), \quad \forall x, x' \in X.$$

Lipschitz constant: denoted as $\|f\|_L$.

Mathematical background

Lipschitz continuity

$f : X \to Y$ is *L-Lipschitz continuous*:

$$d_Y(f(x), f(x')) \leq L d_X(x, x'), \quad \forall x, x' \in X.$$

Lipschitz constant: denoted as $\|f\|_L$.

Hölder continuity

$f : X \to Y$ is α, *L-Hölder continuous* for some $\alpha \leq 1$:

$$d_Y(f(x), f(x')) \leq L d_X(x, x')^\alpha, \quad \forall x, x' \in X.$$

$\alpha < \beta \implies \mathcal{HC}(\beta) \subset \mathcal{HC}(\alpha)$.

Mathematical background 2

Wasserstein distance

For probability measures μ, ν on Ω, the *Wasserstein* distance is defined as:

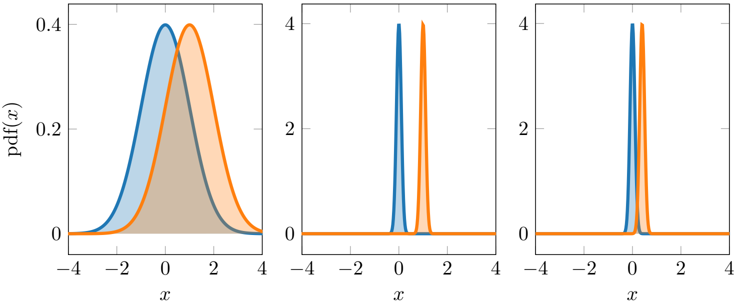
$$\mathcal{W}_1(\mu,\nu) = \sup_{\|f\|_L \leq 1} \left| \int_\Omega f(\omega)(\mu - \nu)(d\omega) \right| \quad \forall \mu, \nu \in \Omega$$

Mathematical background 2

Wasserstein distance

For probability measures μ, ν on Ω, the *Wasserstein* distance is defined as:

$$\mathcal{W}_1(\mu,\nu) = \sup_{\|f\|_L \leq 1} \left| \int_\Omega f(\omega)(\mu-\nu)(d\omega) \right| \quad \forall \mu,\nu \in \Omega$$



Lipschitz MDPs

Lipschitz MDP

S and A metric spaces with distances d_S and d_A.

- $|\mathbb{E}[R(s,a)] - \mathbb{E}[R(s',a')]| \leq L_r(d_S(s,s') + d_A(a,a'))$
- $\mathcal{W}_1(p(\cdot|s,a), p(\cdot|s',a')) \leq L_P(d_S(s,s') + d_A(a,a'))$

Lipschitz Policy

- $\mathcal{W}_1(\pi(\cdot|s), \pi(\cdot|s')) \leq L_\pi d_S(s,s')$.

Lipschitz MDPs

Lipschitz MDP

S and A metric spaces with distances d_S and d_A.

- $|\mathbb{E}[R(s,a)] - \mathbb{E}[R(s',a')]| \leq L_r(d_S(s,s') + d_A(a,a'))$
- $\mathcal{W}_1(p(\cdot|s,a), p(\cdot|s',a')) \leq L_P(d_S(s,s') + d_A(a,a'))$

Lipschitz Policy

- $\mathcal{W}_1(\pi(\cdot|s), \pi(\cdot|s')) \leq L_\pi d_S(s,s')$.

Theorem [Rachelson and Lagoudakis 2010]

Under $\gamma L_P(1 + L_\pi) < 1$, Q^π is Lipschitz-continuous

$$L_{Q^\pi} \leq \frac{L_r}{1 - \gamma L_P(1 + L_\pi)}$$

First results

Performance bound on BC

If $Q^{\pi_I}(s, a)$ is $L_{Q^{\pi_I}}$-LC, then:

$$J^{\pi_E} - J^{\pi_I} \leq \frac{L_{Q^{\pi_I}}}{1-\gamma} \mathbb{E}_{s \sim d^{\pi_E}} [\mathcal{W}_1(\pi_I(\cdot|s), \pi_E(\cdot|s))].$$

First results

Performance bound on BC

If $Q^{\pi_I}(s,a)$ is $L_{Q^{\pi_I}}$-LC, then:

$$J^{\pi_E} - J^{\pi_I} \leq \frac{L_{Q^{\pi_I}}}{1-\gamma} \mathbb{E}_{s \sim d^{\pi_E}} [\mathcal{W}_1(\pi_I(\cdot|s), \pi_E(\cdot|s))].$$

Is this a realistic assumption?

$$L_{Q^\pi} \leq \frac{L_r}{1 - \gamma L_p (1 + L_\pi)}. \tag{1}$$

This bound cannot be improved, but...

Hölder continuity of the Q^π function

The $Q^\pi(s, a)$ function is always $\alpha, L - HC$ with

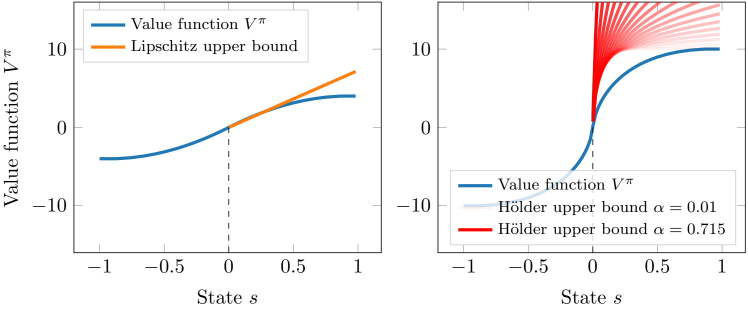
$$0 < \alpha < \overline{\alpha} := \min\left\{1, \frac{-\log\gamma}{\log(L_p(1 + L_\pi))}\right\}.$$

This bound cannot be improved, but...

Hölder continuity of the Q^π function

The $Q^\pi(s, a)$ function is always $\alpha, L - HC$ with

$$0 < \alpha < \overline{\alpha} := \min\left\{1, \frac{-\log\gamma}{\log(L_p(1 + L_\pi))}\right\}.$$



Result over BC

Optimal Error Rate for BC

$$J^{\pi_E} - J^{\pi_I} \leq \frac{L_{Q^{\pi_I},\alpha}}{1-\gamma} \mathbb{E}_{s\sim d^{\pi_E}} \left[\mathcal{W}_1 \left(\pi_E(\cdot|s), \pi_I(\cdot|s) \right)^{\alpha} \right].$$

Best possible performance guarantee using BC:

$$J^{\pi_E} - J^{\pi_I} \leq \mathcal{O}(\varepsilon^{\alpha}),$$

where $\varepsilon^2 = \mathbb{E}_{s\sim d^{\pi_E}}[\|\pi_E(s) - \pi_I(s)\|_2^2]$.

Noise injection

Noisy expert:

$$\forall t \in \mathbb{N} : \qquad \begin{cases} a_{t,E} \sim \pi_E(\cdot|s_t) \\ \eta_t \overset{\text{iid}}{\sim} \mathcal{L} \\ a_t = a_{t,E} + \eta_t \end{cases} ,$$

Noise injection

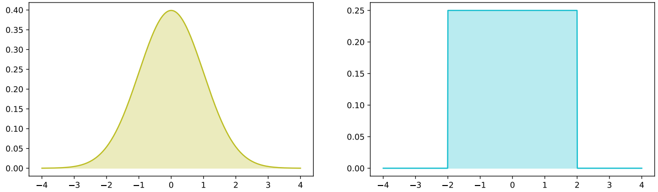
Noisy expert:

$$\forall t \in \mathbb{N}: \quad \begin{cases} a_{t,E} \sim \pi_E(\cdot|s_t) \\ \eta_t \overset{\text{iid}}{\sim} \mathcal{L} \\ a_t = a_{t,E} + \eta_t \end{cases},$$

Noise condition

denoting with TV$(\cdot,\cdot)$ the total variation distance

$$\text{TV}(\mathcal{L}(\cdot + h), \mathcal{L}(\cdot)) \leq L_\ell \|h\|_2, \qquad \forall h \in \mathbb{R}^n.$$



Result of noise injection

Error bound for noise injection

Given noisy expert $\pi_{E,\ell}$ and a noisy imitator $\pi_{I,\ell}$.

$$J^{\pi_{E,\ell}} - J^{\pi_{I,\ell}} \leq \frac{2L_\ell Q_{\max}}{1-\gamma} \mathbb{E}_{s\sim d^{\pi_{E,\ell}}}[\mathcal{W}_1(\pi_E(\cdot|s), \pi_I(\cdot|s))].$$

Note that $Q_{\max} \leq \frac{R_{max}}{1-\gamma}$.

Result of noise injection

Error bound for noise injection

Given noisy expert $\pi_{E,\ell}$ and a noisy imitator $\pi_{I,\ell}$.

$$J^{\pi_{E,\ell}} - J^{\pi_{I,\ell}} \leq \frac{2L_\ell Q_{\max}}{1-\gamma} \mathbb{E}_{s \sim d^{\pi_{E,\ell}}}[\mathcal{W}_1(\pi_E(\cdot|s), \pi_I(\cdot|s))].$$

Note that $Q_{\max} \leq \frac{R_{max}}{1-\gamma}$.

Performance guarantee using BC:

$$J^{\pi_{E,\ell}} - J^{\pi_{I,\ell}} \leq \mathcal{O}(\varepsilon),$$

where $\varepsilon^2 = \mathbb{E}_{s \sim d^{\pi_E}}[\|\pi_E(s) - \pi_I(s)\|_2^2]$.

Result of noise injection

Error bound for noise injection

Given noisy expert $\pi_{E,\ell}$ and a noisy imitator $\pi_{I,\ell}$.

$$J^{\pi_{E,\ell}} - J^{\pi_{I,\ell}} \leq \frac{2L_\ell Q_{\max}}{1-\gamma} \mathbb{E}_{s \sim d^{\pi_{E,\ell}}}[\mathcal{W}_1(\pi_E(\cdot|s), \pi_I(\cdot|s))].$$

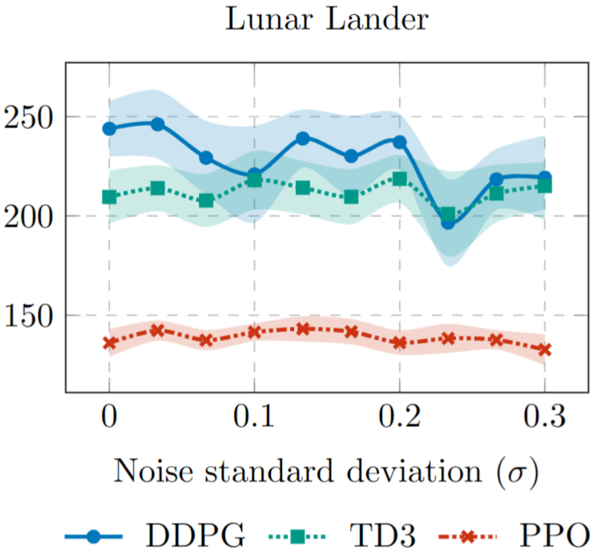
Note that $Q_{\max} \leq \frac{R_{max}}{1-\gamma}$.
Performance guarantee using BC:

$$J^{\pi_{E,\ell}} - J^{\pi_{I,\ell}} \leq \mathcal{O}(\varepsilon),$$

where $\varepsilon^2 = \mathbb{E}_{s \sim d^{\pi_E}}[\|\pi_E(s) - \pi_I(s)\|_2^2]$.

Noise gap

Is $J^{\pi_E} - J^{\pi_{E,\ell}}$ big?


Lunar Lander
250
200
150
0
0.1
0.2
0.3
Noise standard deviation (σ)
DDPG
TD3
PPO

THANK YOU FOR

THE ATTENTION