Generalized Emphatic TD Learning: Bias-Variance Analysis

Hallak,Tamar,Munos,Mannor (2015)

Specific case of ETD(0,β)

Greta Laage

$ETD(0)$ from Sutton, Mahmood and White(2015)

Emphasis

Decaying trace of the importance sampling ratios

$$F_t = \gamma \rho_{t-1} F_{t-1} + 1$$

$$F_0 = 1$$

Algorithm update

$$\begin{aligned}
\theta_{t+1} &= \theta_t + \alpha \rho_t F_t \left(R_{t+1} + \gamma \theta_t^T \phi_{t+1} - \theta_t^T \phi_t \right) \phi_t \\
&= \theta_t + \alpha \left(\rho_t F_t \phi_t R_{t+1} - \rho_t F_t \phi_t (\phi_t - \gamma \phi_{t+1})^T \theta_t \right)
\end{aligned}$$

$ETD(0, \beta)$ from Hallak,Tamar,Munos,Mannor (2015)

Emphasis

$$F_t = \beta \rho_{t-1} F_{t-1} + 1$$
$$F_0 = 1$$

$\beta \in (0, 1)$: parameter that controls the decay rate

Algorithm update

$$\begin{aligned}
\theta_{t+1} &= \theta_t + \alpha \rho_t F_t \left(R_{t+1} + \gamma \theta_t^T \phi_{t+1} - \theta_t^T \phi_t \right) \phi_t \\
&= \theta_t + \alpha \left(\rho_t F_t \phi_t R_{t+1} - \rho_t F_t \phi_t (\phi_t - \gamma \phi_{t+1})^T \theta_t \right)
\end{aligned}$$

Particular values

$\beta = \gamma$: $ETD(0)$
$\beta = 0$: standard TD in off-policy setting
$\beta = 1$: full importance-sampling TD from Precup, Sutton, Dasgupta (2001)

$ETD(0, \beta)$

Emphatic weight vector

$$f(s) = d_\mu(s) \lim_{t->\infty} \mathbb{E}_\mu[F_t | S_t = s]$$

Expected number of time that would be spent in s starting from d_μ.
$d_\mu +$ where you would get to after one step, after two steps, etc.

$$f^T = d_\mu^T (I - \beta P_\pi)^{-1}$$

Convergence

$ETD(0, \beta)$ converges to solution of the projected fixed point equation, where Π_f is the projection to $\Phi, F : \ V = \Pi_f TV$

$ETD(0, \beta)$

Emphatic weight vector

$$f(s) = d_\mu(s) \lim_{t->\infty} \mathbb{E}_\mu[F_t | S_t = s]$$

Expected number of time that would be spent in s starting from d_μ.
$d_\mu +$ where you would get to after one step, after two steps, etc.

$$f^T = d_\mu^T (I - \beta P_\pi)^{-1}$$

Convergence

$ETD(0, \beta)$ converges to solution of the projected fixed point equation, where Π_f is the projection to Φ, F : $V = \Pi_f TV$

Bias (asymptotic error) bound

Contraction property of $\Pi_f T =>$ bias bound. Do we have it ?

$$\begin{aligned} ||\Pi_f Tv_1 - \Pi_f Tv_2||_f &\leq ||Tv_1 - Tv_2||_f \quad (\Pi_j \text{ non-expansion}) \\ &\leq \gamma^2 ||P_\pi(v_1 - v_2)||_f \quad \rightarrow \text{ About } ||P_\pi v||? \end{aligned}$$

Bias of $ETD(0, \beta)$: Contraction property of $\Pi_f T$

We have

$$\begin{aligned}
v^T P_\pi^T F P_\pi v &= \sum_s f(s) (\sum_{s'} P_\pi(s'|s) v(s'))^2 \\
&\leq \sum_s f(s) \sum_{s'} P_\pi(s'|s) v^2(s') \qquad \text{(Jensen's inequality)} \\
&\leq \sum_{s'} v^2(s') \sum_s f(s) P_\pi(s'|s) \\
&\leq v^T diag(f^T P_\pi) v
\end{aligned}$$

Bias of $ETD(0, \beta)$: Contraction property of $\Pi_f T$

We have

$$
\begin{aligned}
v^T P_\pi^T F P_\pi v &= \sum_s f(s) (\sum_{s'} P_\pi(s'|s) v(s'))^2 \\
&\leq \sum_s f(s) \sum_{s'} P_\pi(s'|s) v^2(s') \qquad \text{(Jensen's inequality)} \\
&\leq \sum_{s'} v^2(s') \sum_s f(s) P_\pi(s'|s) \\
&\leq v^T diag(f^T P_\pi) v
\end{aligned}
$$

Then

$$
\begin{aligned}
||v||_f^2 - \beta ||P_\pi v||_f^2 &= v^T F v - \beta v^T P_\pi^T F P_\pi v \\
&\geq v^T F v - \beta v^T diag(f^T P_\pi) v \\
&\geq v^T diag(f^T (I - \beta) P_\pi) v \\
&\geq v^T diag(d_\mu) v \\
&\geq ||v||_{d_\mu}^2 = \sum_s d_\mu(s) v^2(s) \\
&\geq \sum_s \kappa f(s) v^2(s) \qquad \text{where } \kappa = \min_s \frac{d_\mu(s)}{f(s)} \\
&\geq \kappa ||v||_f^2 \qquad \text{and } ||P_\pi v||_f^2 \leq \frac{1-\kappa}{\beta} ||v||_f^2
\end{aligned}
$$

Bias of $ETD(0, \beta)$: Contraction property of $\Pi_f T$

Contraction property wrt the f-weighted norm

For $\beta > \gamma^2(1-\kappa)$,

$$||\Pi_f T v_1 - \Pi_f T v_2||_f \leq \sqrt{\frac{\gamma^2}{\beta}(1-\kappa)}||v_1 - v_2||_f$$

Bias of $ETD(0, \beta)$: Contraction property of $\Pi_f T$

Contraction property wrt the f-weighted norm

For $\beta > \gamma^2(1-\kappa)$,

$$||\Pi_f T v_1 - \Pi_f T v_2||_f \leq \sqrt{\frac{\gamma^2}{\beta}(1-\kappa)}||v_1 - v_2||_f$$

Lemma from Bertsekas and Tsitsiklis(1996):
x^* solution of $x = Ax + b$ ie V^π and $\bar{x}$ solution of $x = \Pi(Ax+b)$ ie V_θ, then $||x^* - \bar{x}|| \leq \frac{1}{1-||\Pi A||}||x^* - \Pi x^*||$

Bias of $ETD(0, \beta)$: Contraction property of $\Pi_f T$

Contraction property wrt the f-weighted norm

For $\beta > \gamma^2(1-\kappa)$,

$$||\Pi_f T v_1 - \Pi_f T v_2||_f \leq \sqrt{\frac{\gamma^2}{\beta}(1-\kappa)}||v_1 - v_2||_f$$

Lemma from Bertsekas and Tsitsiklis(1996):
x^* solution of $x = Ax + b$ ie V^π and $\bar{x}$ solution of $x = \Pi(Ax+b)$ ie V_θ, then $||x^* - \bar{x}|| \leq \frac{1}{1-||\Pi A||}||x^* - \Pi x^*||$

Error bound

$$||\Phi^T\theta^* - V^\pi||_f \leq \frac{1}{\sqrt{1-\frac{\gamma^2}{\beta}(1-\kappa)}}||\Pi_f V^\pi - V^\pi||_f$$

$$||\Phi^T\theta^* - V^\pi||_{d_\mu} \leq \frac{1}{\sqrt{\gamma(1-\frac{\gamma^2}{\beta}(1-\kappa))}}||\Pi_f V^\pi - V^\pi||_f$$

Variance of $ETD(0, \beta)$

$$\theta_{t+1} = \theta_t + \alpha \rho_t F_t \left(R_{t+1} + \gamma \theta_t^T \phi_{t+1} - \theta_t^T \phi_t \right) \phi_t$$

F_t amplitude affects stability => analysis of its variance

Variance bound

$$\mathbb{E}_\mu[Var[F_t | S_t = s]] \leq \frac{\beta^2}{1-\beta} \left(2 + \frac{(1+\beta)||\tilde{P}_{\mu,\pi}||_\infty}{1 - \beta^2 ||\tilde{P}_{\mu,\pi}||_\infty} \right)$$

Where $[\tilde{P}_{\mu,\pi}]_{\bar{s}s} = \sum_a p(s|\bar{s}, \bar{a}) \frac{\pi^2(a|\bar{s})}{\mu(a|\bar{s})}$ the mismatch matrix

Bias-Variance Trade-Off

Variance bound

$$\mathbb{E}_{\mu}[Var[F_t|S_t = s]] \leq \frac{\beta^2}{1-\beta}\left(2 + \frac{(1+\beta)||\tilde{P}_{\mu,\pi}||_{\infty}}{1-\beta^2||\tilde{P}_{\mu,\pi}||_{\infty}}\right)$$

Bias bound

$$||\Phi^T\theta^* - V^{\pi}||_{d_{\mu}} \leq \frac{1}{\sqrt{\gamma(1-\frac{\gamma^2}{\beta}(1-\kappa))}}||\Pi_f V^{\pi} - V^{\pi}||_f$$

β: trade-off parameter

- $\beta \nearrow$, low bias, high variance
- $\beta \searrow$, high bias, low variance