Actor-critic + CS

Friday, May 22, 2020 5:25 AM

1) infinite-horizon optimal control problem ??

2) (Hamilton-Jacobi-Bellman eqn)??

(Infinite-horizon optimal control)

control problem → cost function J (d.o.f. state x and other variables)

$J = f(x(t), \ldots)$

optimal control → set of DE that min. cost fn

example:-

cost function

$$J = \phi[x(t_0), t_0, x(t_f), t_f] + \int_{t_0}^{t_f} L[\;]$$

$$J = \frac{1}{2} x^T S_f x + \frac{1}{2}\int_{t_0}^{t_f} \left((x^T Q x) + (u^T R u) \right) dt$$

subject to t_0; DE $\dot{x} = Ax + Bu$; $x(t_0) = x_0$

$Q, R \to$ +ve definite as J should be +ve

LQR → linear Quadratic Control.

finite time horizon

infi. time horizon: $t_0 = 0$, $t_f \to \infty$; A, B, Q & $R \to$ constant, so that cost should be +ve.

$Q, R \to$ +ve semi definite, +ve definite

$u = -K(t)x(t)$

$K(t) = R^{-1} B^T S(t)$

$\dot{S}(t) = -S(t)A - A^T S(t) + S(t) B R^{-1} B^T S(t) - Q$

infinite horizon ⇒

$0 = -SA + A^T S + SBR^{-1}B^T S - Q$

3) Actor → optimal control (policy)

Critic → optimal value fuction (state)

identifier → estimates the systems dynamic online

DNN → Hop-field type component and novel RISE (Robust Integral sign of error) learns the system dynamics

Actor-critic Architecture:-

$\dot{x} = f(x, u)$; $x(t) \in \mathcal{X} \subseteq \mathbb{R}^n$ states

$u(t) \in \mathcal{U} \subseteq \mathbb{R}^m$ control input.

optimal cost is given as

$$V^*(x(t)) = \min_{u(t) \in \psi(\mathcal{U})} \int_t^\infty r(x(s), u(x(s)))\, ds$$

Critic

$r(x, u) \to$ reward / local cost or immediate cost

$$r(x, u) = Q(x) + u^T R u$$

$$\dot{x} = f(x) + g(x) u$$

$$u^*(x) = -\frac{1}{2} R^{-1} g^T(x) \frac{\partial V^*(x)}{\partial x}^T$$

optimal control. Actor

Hamiltonian of the system

condition for the system

$$H(x, u, V_x) = V_x F_u + r_u$$

$V_x \to$ gradient of Value function.

$\frac{\partial V}{\partial x}$

$F_u = f(x) + g(x) u$

system dynamics with control $u(x)$

$r_u = r(x, u) \leftarrow$ reward / local cost

$V^*(x)$ & $u^*(x)$ satisfy

$$\boxed{H(x, u^*, V_x^*) = V_x^* F_{u^*} + r_{u^*} = 0}$$

we use approximators $\hat{u}$, $\hat{V}$, $\hat{F}$

we use approximators ($\hat{u}$, $\hat{V}$, $\hat{F}$)

$\delta_{hjb} \rightarrow$ Bellman error

$\hookrightarrow \hat{H} - H^* = H(x, \hat{x}, \hat{u}, \hat{V}_x) - \cancel{H(x, u^*, V_x^*)}$ 0

$$\delta_{hjb} = H(x, \hat{x}, \hat{u}, \hat{V}_x)$$
$$= \hat{V}_x \hat{F}_{\hat{u}} + r(x, \hat{u})$$

actor and critic network are learns by δ_{hjb}

and DNN using identifn error

$$\boxed{\Delta x = \tilde{x}(t) = x(t) - \hat{x}(t)}$$

Actor - critic design)

Actor $\rightarrow$ optimal control policy

$$u^*(x) = -\frac{1}{2} R^{-1} g^T(x) \left(\phi'(x)^T W + \varepsilon_V'(x)^T\right)$$

critic $\rightarrow$ optimal value function

$$V^*(x) = W^T \phi(x) + \varepsilon_V(x)$$

Seperate NN used for actor and critic.

Can be overruled;

Two different types of updates is given

critic $\leftrightarrow$ Least square error

Actor $\leftrightarrow$ squared bellman error

~~However~~, However, we can do that simply ??

logarithamic error ??

Is it important to use identifier ??