Értékelőfüggvény közelítése a megerősítéses tanulásban

Kaczúr Flórián

Témavezető: Csáji Balázs Csanád

2023. június 2.

Markov-döntési folyamatok

Markov-döntési folyamatok

- $X = \{1, \ldots, n\}$ állapottér

Markov-döntési folyamatok

- $X = \{1, \ldots, n\}$ állapottér
- A véges akciótér, $\mathcal{A} : X \to \mathcal{P}(A)$

Markov-döntési folyamatok

- $X = \{1, \ldots, n\}$ állapottér
- A véges akciótér, $\mathcal{A} : X \to \mathcal{P}(A)$
- $p_{ij}(a)$ átmenetvalószínűség

Markov-döntési folyamatok

- $X = \{1, \ldots, n\}$ állapottér
- A véges akciótér, $\mathcal{A} : X \to \mathcal{P}(A)$
- $p_{ij}(a)$ átmenetvalószínűség
- $g(i, j)$ költség/jutalom

Markov-döntési folyamatok

- $X = \{1, \ldots, n\}$ állapottér
- A véges akciótér, $\mathcal{A} : X \to \mathcal{P}(A)$
- $p_{ij}(a)$ átmenetvalószínűség
- $g(i, j)$ költség/jutalom

Politika: $\mu : X \to \Delta(A)$

Markov-döntési folyamatok

- $X = \{1, \ldots, n\}$ állapottér
- A véges akciótér, $\mathcal{A} : X \to \mathcal{P}(A)$
- $p_{ij}(a)$ átmenetvalószínűség
- $g(i, j)$ költség/jutalom

Politika: $\mu : X \to \Delta(A)$

Átmenetmátrix: $p_{ij} = \sum_{a \in \mathcal{A}(i)} p_{ij}(a) \cdot \mu(i, a)$, feltétel: *irreducibilitás*

Markov-döntési folyamatok

- $X = \{1, \ldots, n\}$ állapottér
- A véges akciótér, $\mathcal{A} : X \to \mathcal{P}(A)$
- $p_{ij}(a)$ átmenetvalószínűség
- $g(i, j)$ költség/jutalom

Politika: $\mu : X \to \Delta(A)$

Átmenetmátrix: $p_{ij} = \sum_{a \in \mathcal{A}(i)} p_{ij}(a) \cdot \mu(i, a)$, feltétel: *irreducibilitás*

Értékelőfüggvény:

$$J^{\mu}(i) = \limsup_{N \to \infty} \mathbb{E}\left\{ \sum_{k=0}^{N-1} \alpha^k g\left(i_k, i_{k+1}\right) | i_0 = i \right\}$$

Megerősítéses tanulás

Megerősítéses tanulás

Továbbiakban: J^{μ} közelítése.

Megerősítéses tanulás

Továbbiakban: J^{μ} közelítése.

Gond: P-t és g-t nem ismerjük.

Megerősítéses tanulás

Továbbiakban: J^{μ} közelítése.

Gond: P-t és g-t nem ismerjük.

Számításigényes módszerek.

Projektált Bellman-módszerek

$span\{\Phi_1, \ldots, \Phi_m\} \leq \mathbb{R}^n$

Projektált Bellman-módszerek

$span\{\Phi_1, \ldots, \Phi_m\} \leq \mathbb{R}^n$

Fixpont-egyenlet:

$$\Pi_\xi^\Phi T^{(\lambda)}\left(\Phi r_\lambda^*\right) = \Phi r_\lambda^*,$$

Projektált Bellman-módszerek

$span\{\Phi_1, \ldots, \Phi_m\} \leq \mathbb{R}^n$

Fixpont-egyenlet:

$$\Pi_\xi^\Phi T^{(\lambda)}\left(\Phi r_\lambda^*\right) = \Phi r_\lambda^*,$$

ahol

- $P^T \xi = \xi$
- $T^{(\lambda)} = (1-\lambda)\sum_{\ell=0}^{\infty} \lambda^\ell T^{\ell+1}$, $0 \leq \lambda < 1$
- $(TJ)(i) = \sum_{j=1}^{n} p_{ij} \cdot \left(g(i,j) + \alpha J(j)\right), \quad i = 1, \ldots, n$
- $\Phi \in \mathbb{R}^{n \times m}$.

Becslés

Állítás

$\Pi_{\xi}^{\Phi} T^{(\lambda)}$ *kontrakció. [1]*

Becslés

Állítás

$\Pi_\xi^\Phi T^{(\lambda)}$ *kontrakció. [1]*

Állítás

$$\|J_\mu - \Phi r_\lambda^*\|_\xi \leq \left(\frac{1}{\sqrt{1-\alpha_\lambda^2}}\right)\|J_\mu - \Pi_\xi^\Phi J_\mu\|_\xi, \text{ ahol } \alpha_\lambda = \frac{\alpha(1-\lambda)}{1-\alpha\lambda}.$$

[1]

Eljárások

$$\Pi_{\xi}^{\Phi} T^{(\lambda)}\left(\Phi r_{\lambda}^{*}\right) = \Phi r_{\lambda}^{*}$$

Eljárások

$$\Pi_{\xi}^{\Phi} T^{(\lambda)}\left(\Phi r_{\lambda}^{*}\right) = \Phi r_{\lambda}^{*} \qquad \rightarrow \qquad A\, r_{\lambda}^{*} = b$$

Eljárások

$$\Pi_{\xi}^{\Phi} T^{(\lambda)}\left(\Phi r_{\lambda}^{*}\right) = \Phi r_{\lambda}^{*} \qquad \rightarrow \qquad A\, r_{\lambda}^{*} = b$$

$$\downarrow \qquad \downarrow$$

$$A_k \qquad b_k$$

Eljárások

$$\Pi_{\xi}^{\Phi} T^{(\lambda)}\left(\Phi r_{\lambda}^{*}\right) = \Phi r_{\lambda}^{*} \qquad \rightarrow \qquad \begin{array}{cc} A\, r_{\lambda}^{*} = b \\ \downarrow \quad \downarrow \\ A_k \quad b_k \end{array}$$

1. LSTD(λ): $\hat{r}_k = A_k^{-1} b_k$

Eljárások

$$\Pi_{\xi}^{\Phi} T^{(\lambda)}\left(\Phi r_{\lambda}^{*}\right) = \Phi r_{\lambda}^{*} \qquad \rightarrow \qquad \begin{array}{cc} A\, r_{\lambda}^{*} = b \\ \downarrow \quad \downarrow \\ A_k \quad b_k \end{array}$$

1. LSTD(λ): $\hat{r}_k = A_k^{-1} b_k$
2. LSPE(λ): $r_{k+1} = r_k - \gamma G_k \left(A_k r_k - b_k\right)$

Eljárások

$$\Pi_\xi^\Phi T^{(\lambda)}\left(\Phi r_\lambda^*\right) = \Phi r_\lambda^* \qquad \rightarrow \qquad \begin{array}{cc} A\, r_\lambda^* = b \\ \downarrow \quad \downarrow \\ A_k \quad b_k \end{array}$$

1. LSTD(λ): $\hat{r}_k = A_k^{-1} b_k$

2. LSPE(λ): $r_{k+1} = r_k - \gamma G_k \left(A_k r_k - b_k\right)$

3. TD(λ): $r_{k+1} = r_k - \gamma_k z_k q_{k,k}$, ahol

$$q_{k,k} = \Phi(i_k)^T r_k - \alpha \Phi(i_{k+1})^T r_k - g(i_k, i_{k+1}),$$

$$z_k = \sum_{h=0}^{k} (\alpha\lambda)^{k-h} \Phi(i_h).$$

Felfedezés javítása

- kezdeti ζ eloszlás

Felfedezés javítása

- kezdeti ζ eloszlás
- $c_{\ell,s}(r_k) = \alpha^{N_s-\ell}\Phi(i_{N_s-1,s})^T r_k + \sum_{h=\ell}^{N_s-1} \alpha^{h-\ell} g\left(i_{h,s}, i_{h+1,s}\right)$

Felfedezés javítása

- kezdeti ζ eloszlás
- $c_{\ell,s}(r_k) = \alpha^{N_s-\ell}\Phi(i_{N_s-1,s})^T r_k + \sum_{h=\ell}^{N_s-1} \alpha^{h-\ell} g\left(i_{h,s}, i_{h+1,s}\right)$
- $r_{k+1} = \arg\min_{r\in\mathbb{R}^m} \sum_{s=0}^{T} \sum_{\ell=0}^{N_s-1} \left(\Phi(i_{\ell,s})^T r - c_{\ell,s}(r_k)\right)^2$

Felfedezés javítása

- kezdeti ζ eloszlás
- $c_{\ell,s}(r_k) = \alpha^{N_s-\ell}\Phi(i_{N_s-1,s})^T r_k + \sum_{h=\ell}^{N_s-1} \alpha^{h-\ell} g\left(i_{h,s}, i_{h+1,s}\right)$
- $r_{k+1} = \arg\min_{r\in\mathbb{R}^m} \sum_{s=0}^{T} \sum_{\ell=0}^{N_s-1} \left(\Phi(i_{\ell,s})^T r - c_{\ell,s}(r_k)\right)^2$
- $\Phi r = \Pi_\zeta^\Phi T^{(\lambda)} \Phi r.$

Hivatkozások

Bertsekas, Dimitri P. "Dynamic programming and optimal control 3rd edition, volume ii." Belmont, MA: Athena Scientific (2011).