IA 718 Tópicos em Sistemas Inteligentes

3-Introdução ao Processo de Decisão de Markov

Conteúdo

1. Equações de otimalidade
2. Problemas com horizonte finito
3. Problemas com horizonte infinito
4. Iteração de valor (*value iteration*)
5. Iteração de estratégia (*policy iteration*)
6. Iteração híbrida
7. Programação linear

Notação

$S = \{1, 2, ..., \|S\| \}$	espaço de estado
A	conjunto de ações/decisões
X_t	conjunto decisões factíveis
x_t	variável de decisão genérica
$x_t \in X_t$	
$p_t(S_{t+1} \mid S_t, x_t)$	matriz transição de estado

Processos de Markov

- Processo estocástico

  $\{S_t\}$, $t = 1, 2, ...$ em geral $t \in \mathsf{T}$

  S_t variável aleatória

  exemplo nível semanal de um estoque

- Propriedade Markoviana

$$P(S_{t+1} = j \mid S_0 = k_0, S_1 = k_0, \ldots, S_{t-1} = k_{t-1}, S_t = j\} = P(S_{t+1} = j \mid S_t = j\}$$

$t = 0,1,2,\ldots$ e qualquer $i, j, k_0, k_1, \ldots, k_{t-1}$

$P(S_{t+1} = j \mid S_t = j\}$ probabilidades transição

$P(S_{t+1} = j \mid S_t = j\} = P(S_1 = j \mid S_0 = j\} = p_{ij} \;; \quad \forall t$ estacionárias

$P(S_{t+n} = j \mid S_t = j\} = P(S_n = j \mid S_0 = j\} = p_{ij}^n \;; \quad \forall t$ transição n passos

$$p_{ij}^n \geq 0, \quad \sum_{j=0}^{|S|} p_{ij}^n = 1, \; \forall i; \; n = 0,1,2,\ldots$$

1-Equações de otimalidade

$$\max_{\pi \in \Pi} E\left\{ \sum_{t=0}^{T} \gamma^t C_t^{\pi}(S_t, X_t^{\pi}(S_t)) \right\}$$

C_t^{π}	custo imediato
S_t	estado
X_t^{π}	função decisão
Π	conjunto funções de decisão
γ	fator de esquecimento
$S^M(S_t, x_t, W_{t+1})$	transição de estado

- Equação de Bellman: caso determinístico

$$x_t^*(S_t) = arg \max_{x_t \in X_t} (C_t(S_t, x_t) + \gamma V_{t+1}(S_{t+1}))$$

$$S_{t+1} = S^M(S_t, x_t)$$

$$V_t(S_t) = \max_{x_t \in X_t} (C_t(S_t, x_t) + \gamma V_{t+1}(S_{t+1}(S_t, x_t)))$$

$$= C_t(S_t, x_t^*) + \gamma V_{t+1}(S_{t+1}(S_t, x_t^*))$$

- Equação de Bellman: caso estocástico

  – exemplo controle de estoque

S_t	estoque em t
x_t	reposição em t
$\hat{D}_{t+1}$	demanda entre t e $t + 1$

$$S_{t+1}(S_t, x_t) = max\{0, S_t + x_t - \hat{D}_{t+1}\}$$

$\hat{D}_{t+1}$ variável aleatória em t

$$P^D(d) = P[\hat{D} = d]$$

$$Prob(S_{t+1} = s') = \begin{cases} 0 & s' > S_t + x_t \\ P^D(S_t + x_t - s') & 0 < s' \leq S_t + x_t \\ \sum_{d=S_t+x_t}^{\infty} P^D(d) & s' = 0 \end{cases}$$

- Equação de Bellman: caso estocástico

$$V_t(S_t) = \max_{x_t \in X_t} \left(C_t(S_t, x_t) + \gamma \sum_{s' \in S} P(S_{t+1} = s' \mid S_t, x_t) V_{t+1}(s')\right) \qquad (3.3)$$

$$V_t(S_t) = \max_{x_t \in X_t} \left(C_t(S_t, x_t) + \gamma E\{V_{t+1}(S_{t+1}(S_t, x_t)) \mid S_t\}\right) \qquad (3.4)$$

$P(S_{t+1} \mid S_t, x_t)$ probabilidade de S_{t+1} dado S_t e x_t

$$V_t(S_t) = \max_{x_t \in X_t} \left(C_t(S_t, x_t) + \gamma E\{V_{t+1}(S_{t+1}) \mid S_t\}\right)$$ forma compacta de (3.4)

$$S_{t+1} = S^M(S_t, x_t, W_{t+1})$$

$S_t = s,\ S_{t+1} = s'$

$p_{ss'}(x) = P(S_{t+1} = s' \mid S_t = s, x_t = x)$

Se $x = X_t^{\pi}(s)$ então

$p_{ss'}^{\pi}(x) = P(S_{t+1} = s' \mid S_t = s, X_t^{\pi}(s) = x)$

$P_t^{\pi} = [p_{ss'}^{\pi}]$ matriz transição sob estratégia π

$c_t^{\pi}(s) = [C_t(s, X_t^{\pi}(s))]$ vetor coluna

$v_{t+1}(s) = [V_{t+1}(s)]$ vetor coluna

Então (3.3) é equivalente à

$$\begin{bmatrix} \vdots \\ v_t(s) \\ \vdots \end{bmatrix} = \max_{\pi} \left(\begin{bmatrix} \vdots \\ c_t^{\pi}(s) \\ \vdots \end{bmatrix} + \begin{bmatrix} \vdots \\ p_{ss'}^{\pi} \\ \vdots \end{bmatrix} \begin{bmatrix} \vdots \\ v_{t+1}(s') \\ \vdots \end{bmatrix} \right) \tag{3.7}$$

$$v_t = \max_{\pi} \left(c_t^{\pi} + \gamma P_t^{\pi} v_{t+1} \right) \tag{3.8}$$

- Equação (3.8) é resolvida determinando x_t para cada estado s

- Resultado é o vetor

$$x_t^* = (x_t^*(s))_{s \in S}$$

- Equivale a determinar a melhor estratégia (política ótima)

- Equação de Bellman e a função objetivo original

$$\max_{\pi \in \Pi} E\left\{ \sum_{t=0}^{T} \gamma^t C_t^{\pi}(S_t, X_t^{\pi}(S_t)) \right\}$$

- Valor esperado da função objetivo a partir de t utilizando estratégia π

$$F_t^{\pi}(S_t) = E\left\{ \sum_{t'=t}^{T-1} C_{t'}(S_{t'}, X_{t'}^{\pi}(S_{t'})) + C_T(S_T) \mid S_t \right\}$$

- Equação de otimalidade

$$V_t^{\pi}(S_t) = C_t(S_t, X_t^{\pi}(S_t)) + E\left\{ V_{t+1}^{\pi}(S_{t+1}) \mid S_t \right\}$$

- Resolvendo equação de otimalidade recursivamente

$$F_t^{\pi}(S_t) = V_t^{\pi}(S_t)$$

$$F^* = \max_{\pi \in \Pi} F_t^{\pi}(S_t) = V(S_t) \qquad (3.9)$$

- $V(S_t)$ é uma solução de (3.4) / (3.3)
- (3.9) mostra equivalência entre
  - o valor do estado S_t e prosseguir utilizando a estratégia ótima
  - o valor ótimo da função quando no estado S_t

- Matriz de transição de estado

P_t^{π} ?

– conhecida

– calculada a partir de $S_{t+1} = S^M (S_t, x_t, W_{t+1})$

$$1_{(X)} = \begin{cases} 1 & X \text{ é verdadeiro} \\ 0 & \text{caso contrário} \end{cases}$$

$$P_t(S_{t+1} \mid S_t, x_t) = E 1_{(s' = S^M (S_t, x_t, W_{t+1}))} = \sum_{\omega_{t+1} \in \Omega_{t+1}} P(\omega_{t+1}) 1_{(s' = S^M (S_t, x_t, \omega_{t+1}))}$$

- Custos estocásticos

$$\hat{C}_{t+1}(S_t, x_t, W_{t+1})$$

$$V_t(S_t) = \max_{x_t} E\left\{\hat{C}_{t+1}(S_t, x_t, W_{t+1}) + \gamma V_{t+1}(S_{t+1}) \mid S_t\right\}$$

$$\hat{C}_t(S_t, x_t) = E\left\{\hat{C}_{t+1}(S_t, x_t, W_{t+1}) \mid S_t\right\}$$

- Equação de Bellman e a notação operador *M*

$$v_t = \max_{\pi}\left(c_t^{\pi} + \gamma P_t^{\pi} v_{t+1}\right) \qquad (3.8)$$

$M = max\ (min)$ em (3.8)

$$Mv(s) = \max_{\pi}(C_t(s,x) + \gamma \sum_{s' \in s} P_t(s' \mid s,x) v_{t+1}(s')) \qquad \text{componente vetor}$$

$$Mv = \max_{\pi}(c_t^{\pi} + \gamma P_t^{\pi} v_{t+1}) \qquad \text{vetor}$$

$$M : V \rightarrow V$$

$$M^{\pi}(v) = c_t^{\pi} + \gamma P^{\pi} v$$

2-Problemas com horizonte finito

inicializar $V_T(S_t)$

$t = T - 1$

para todo $S_t \in S$ calcular

$$V_t(S_t) = \max_{x_t \in X_t} (C_t(S_t, x_t) + \gamma \sum_{s' \in S} P(s' \mid S_t, x_t) V_{t+1}(s'))$$

$t = t - 1$

até que $t = 0$

3-Problemas com horizonte infinito

- Dados do problema constantes
  - parâmetros
  - custos
  - função transição

- Sistema em estado estacionário

$$V_t(S_t) = \max_{x_t} E\{C_{t+1}(S_t, x_t, W_{t+1}) + \gamma V_{t+1}(S_{t+1}) \mid S_t\}$$

$$V(s) = lim_{t \to \infty} V_t(S_t)$$

$$V(s) \equiv \max_{\pi \in \Pi} E\left\{ \sum_{t=0}^{\infty} \gamma^t C_t(S_t, X_t^{\pi}(S_t)) \right\}$$

$P^{\pi,t} = \prod_{t'=0}^{t-1} P_{t'}^{\pi}, \; P^{\pi,0} = I$ matriz transição t passos

$v_t^{\pi} = \sum_{t'=t}^{\infty} \gamma^{t'-t} P^{\pi,t'-t} c_{t'}^{\pi}$ valor associado à π

$\pi_0, \pi_1 = \pi_2 = \cdots = \pi$

$v^{\pi_0} = c^{\pi_0} + \sum_{t'=1}^{\infty} \gamma^{t'} P^{\pi,t'-1} c_{t'}^{\pi}$

$$v^{\pi_0} = c^{\pi_0} + \sum_{t'=1}^{\infty} \gamma^{t'} P^{\pi, t'-1} c_{t'}^{\pi}$$

$$v^{\pi_0} = c^{\pi_0} + \sum_{t'=1}^{\infty} \gamma^{t'} \left(\prod_{t''}^{t'} P_{t''}^{\pi}\right) c_{t'}^{\pi}$$

$$v^{\pi_0} = c^{\pi_0} + \gamma P^{\pi_0} \sum_{t'=1}^{\infty} \gamma^{t'-1} \left(\prod_{t''}^{t'} P_{t''}^{\pi}\right) c_{t'}^{\pi}$$

$$v^{\pi_0} = c^{\pi_0} + \gamma P^{\pi_0} v^{\pi}$$

Se $\pi_0 = \pi_1 = \pi_2 = \cdots = \pi$ então

$$v^{\pi} = c^{\pi} + \gamma P^{\pi} v^{\pi}$$

$$v^{\pi} = c^{\pi} + \gamma P^{\pi} v^{\pi}$$

$$\Downarrow$$

$$v^{\pi} = (I - \gamma P^{\pi})^{-1} c^{\pi}$$

$$\equiv$$

$$M^{\pi}(v) = c^{\pi} + \gamma P^{\pi} v$$

4-Iteração de valor

inicializar $v^0(s) = 0 \ \forall s \in S$

$\varepsilon > 0$; $n = 0$

repetir

$n = n + 1$

para cada $s \in S$ calcular

$$v^n(s) = \max_{x \in X}(C(s,x) + \gamma \sum_{s' \in S} P(s' \mid s,x) v^{n\text{-}1}(s'))$$

$$\pi^\varepsilon \leftarrow v^\varepsilon = v^n$$

até que $|| v^n - v^{n-1} || < \varepsilon(1 - \gamma)/ 2\gamma$

obs: $||v|| = max_s |v(s)|$

- Algoritmo de Gauss-Seidel

inicializar $v^0(s) = 0 \ \forall s \in S$

$\varepsilon > 0$; $n = 0$

repetir

$n = n + 1$

para cada $s \in S$ calcular

$$v^n(s) = \max_{x \in X}(C(s,x) + \gamma(\sum_{s' < s} P(s' \mid s,x) v^{n-1}(s') + \sum_{s' \geq s} P(s' \mid s,x) v^{n-1}(s'))$$

$\pi^\varepsilon \leftarrow v^\varepsilon = v^n$

até que $|| v^n - v^{n-1} || < \varepsilon(1 - \gamma)/ 2\gamma$

5-Iteração de estratégia

inicializar π^0

$n = 0$

repetir

$n = n + 1$

calcular matriz de transição $(P^{\pi})^{n-1}$

calcular $(c^{\pi})^{n-1}(s) = C(s,(X^{\pi})^{n-1}), \forall s \in S$

resolver $(I - (P^{\pi})^{n-1})\, v = (c^{\pi})^{n-1}$ // fornece solução $(v^{\pi})^n$

determinar $x^n(s) = \underset{x \in X}{arg\,max}(C(x) + \gamma P^{\pi}(v^{\pi})^n) \quad \forall s \in S$

até que $x^{n-1}(s) = x^{n-1}(s), \forall s \in S$

6-Iteração híbrida

inicializar $v^0(s) = v \; \forall s \in S$

$\varepsilon > 0$; $n = 0$; K

repetir

$n = n + 1$

para cada $s \in S$ calcular

$$x^n(s) = \underset{x \in X}{arg\,max}\left(C(s,x) + \gamma \sum_{s' \in S} P(s' \mid s,x) v^{n-1}(s')\right) \quad // \text{ fornece } \pi^n$$

$m = 0$

$u^n(0) = c^\pi + \gamma (P^\pi)^n \, v^{n-1}$

enquanto $k < K$ fazer

$u^n(k+1) = c^\pi + \gamma (P^\pi)^n \, u^n(k)$

$v^n = u^n(K)$

até que $|| \, u^n(0) - v^{n-1} || < \varepsilon(1 - \gamma)/2\gamma$

7-Programação linear

- Se $v(s) \geq \max_{x \in X}(C(s,x) + \gamma \sum_{s' \in S} P(s' \mid s,x) v^{n\text{-}1}(s')), \forall s \in S$ (*)
  - v é um limitante superior de v^*
  - $v^* = c + \gamma P v^*$ é o menor valor de v que satisfaz (*)

$$\min_{v} \sum_{s \in S} \beta_s v(s)$$
$$s.a.\ v(s) \geq C(s,x) + \gamma \sum_{s' \in S} P(s' \mid s,x) v(s'), \quad \forall s,x$$

Observação

Este material refere-se às notas de aula do curso IA 718 Tópicos em Sistemas Inteligentes da Faculdade de Engenharia Elétrica e de Computação da Unicamp. Não substitui o livro texto, as referências recomendadas e nem as aulas expositivas. Este material não pode ser reproduzido sem autorização prévia dos autores. Quando autorizado, seu uso é exclusivo para atividades de ensino e pesquisa em instituições sem fins lucrativos.