Μάθημα 13ο (20/11/18)

~ Δυναμικός Προγραμματισμός ~

Πλαίσιο

- Διακριτός χρόνος
- Πεπερασμένος Χρονικός Ορίζοντας
- Αριθμήσιμος Χώρος καταστάσεων
- Πεπερασμένες το πλήθος διαθέσιμες αποφάσεις σε κάθε στιγμή απόφασης
- Ακολουθιακή λήψη αποφάσεων με γνώση της κατάστασης σε κάθε στιγμή απόφασης

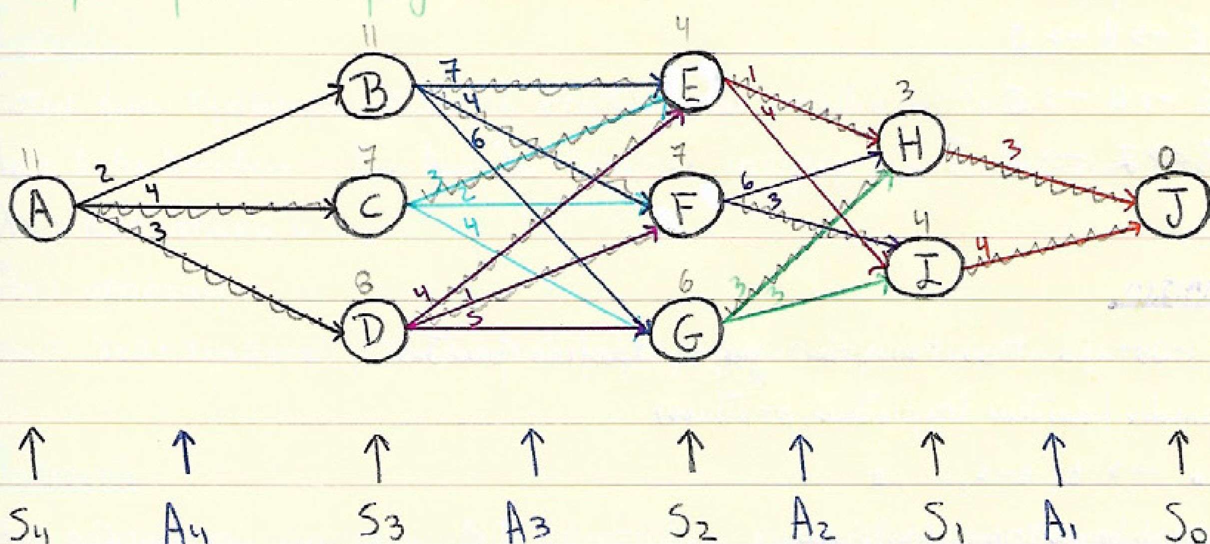
απόφαση

$$X_t \xrightarrow{a_t} X_{t-1} \xrightarrow{a_{t-1}} X_{t-2} \dots \xrightarrow{a_1} X_0$$

κατάσταση   απομένουν t-1 αποφάσεις

- Κάθε λήψη απόφασης συνοδεύεται με κόστος (που εξαρτάται από την τρέχουσα κατάσταση, την απόφαση και το στάδιο)
- Ζητούμενο ελαχιστοποίηση του συνολικού κόστους
- Η επόμενη κατάσταση εξαρτάται από την τρέχουσα κατάσταση και την απόφαση

Το πρόβλημα του αμαξά



Ελαχ. διαδρομή A → J

$c(x \to y)$: κόστος από την x στην y

$U_n(x)$: υπολειπόμενο κόστος από την x στην y

$a_n^*(x)$: βέλτιστη απόφαση στην πόλη

στάδιο

Αρχη Bellman

Αν μια ακολουθία καταστάσεων - αποφάσεων είναι βέλτιστη, τότε το μέρος της από μια ενδιάμεση κατάσταση και μετά είναι επίσης βέλτιστο για το υπολειπόμενο συνολικό κόστος από εκείνη την κατάσταση και μετά.

Λύση του Παραδείγματος με αρχή Bellman

$U_n(x) = \min_y \{c(x \to y) + U_{n-1}(y)\}$ εξίσωση Βελτιστοποίησης

$U_0(J) = 0$

$U_1(H) = 3 + U_0(J) = 3$, $a_1^*(H): H \to J$

$U_1(I) = 4 + U_0(J) = 4$, $a_1^*(I): I \to J$

$U_2(E) = \min(1 + U_1(H), 4 + U_1(I)) = \min(4, 8)$, $a_2^*(E): E \to H$

$U_2(F) = \min(6 + U_1(H), 3 + U_1(I)) = \min(9, 7)$, $a_2^*(F): F \to I$

⋮

Βέλτιστες Διαδρομές

$A \to C \to E \to H \to J$

$A \to D \to E \to H \to J$

$A \to D \to F \to I \to J$

Γενικο Πλαισιο

- Δυναμικό σύστημα που "κινείται" για t χρονικές μονάδες
- t : # χρονικών μονάδων (περιόδων, σταδίων)
- n : στάδιο $\to t, t-1, \dots, 0$
  = # αποφάσεων που απομένουν
- x_n : κατάσταση στο στάδιο n
  (πριν τη λήψη της n-οστής απόφασης)
- a_n : n-οστή απόφαση (απομένουν n-1 αφού ληφθεί)

- $\mathcal{X}$: χώρος καταστάσεων (αριθμήσιμος)
- $\mathcal{A}(x_n, n)$: δέσμη διαθέσιμων αποφάσεων στο στάδιο n, στην κατάσταση n

Δυναμική Συστήματος

$P_{xy}(a;n)$ = (δεσμευμένη) πιθανότητα η επόμενη κατάσταση να είναι η y δεδομένων των: στάδιο n, κατάσταση x, απόφαση a

Δομή κόστους

$c(x, a; n)$: κόστος όταν στο στάδιο n στην κατάσταση x ληφθεί η απόφαση a

$c(x; 0)$: τελικό κόστος στην κατάσταση x

Κριτήριο Βελτιστότητας

$$\min E\left[c(X_t, A_t; t) + c(X_{t-1}, A_{t-1}; t-1) + \dots + c(X_1, A_1; 1) + c(X_0; 0)\right]$$

Βασικές Υποθέσεις ΔΠ

ΔΠ = Ακολουθιακή λήψη αποφάσεων

+

(i) Μαρκοβιανή Δυναμική
(ii) -//- Δομή κόστους

} Μόνο το παρόν αρκεί για το μέλλον (x, a, n)

(iii) Προσθετικότητα όσον αφορά τα κόστη στην αντικειμενική

; → (iv) Τέλεια Παρατήρηση της κατάστασης πριν από κάθε απόφαση

Ιστορία

Έστω ένα πρόβλημα Δ.Π, t σταδίων και στάδιο n. Τότε,

$H_n = X_t A_t X_{t-1} A_{t-1} \cdots X_{n+1} A_{n+1} X_n$ λέγεται ιστορία της διαδικασίας στο στάδιο n

X_n: κατάσταση n

A_n: απόφαση n

$h_n = x_t a_t x_{t-1} a_{t-1} \cdots x_{n+1} a_{n+1} x_n$ μια πραγματοποίηση της H_n

Πολιτική

Μια πολιτική για να είναι Δ.Π t-σταδίων είναι $\Pi = (\pi(t), \pi(t-1), \dots, \pi(1))$ όπου

$\pi(n)$: σύνολο δυνατών ιστοριών $\underbrace{\text{στο στάδιο } n}_{\text{Π.Ο}}$ $\longrightarrow$ σύνολο των κατανομών $\underbrace{\text{στο σύνολο των αποφάσεων}}_{\text{σύνολο καταστάσεων}}$

↑
απεικόνιση

$\Pi_{h_n}(a;n)$ = πιθανότητα να λάβω την απόφαση $a \in A(x_n;n)$ αν είμαι στο στάδιο n και έχω παρατηρήσει την ιστορία $h_n = x_t a_t x_{t-1} a_{t-1} \cdots x_{n+1} a_{n+1} x_n$

$$\Pi(n): h_n(x_t a_t \cdots x_{n+1} a_{n+1} x_n) \longrightarrow \underbrace{(\Pi_{h_n}(a;n):\ a \in A(x_n;n))}_{\text{σύνολο πιθ. στο } A(x_n,n)}$$

Είδη Πολιτικών

1 Ως προς την εξάρτηση από την ιστορία

H (ιστοριοεξαρτ) ⟶ NM

↓

M (μαρκοβιανές) ⟶ NS

↓

S (στάσιμες)

$$\Pi \text{ είναι } M \iff \underset{\substack{\uparrow \\ x_t a_t \cdots x_{n+1} a_{n+1} x_n}}{\Pi_{h_n}(a;n)} = \Pi_{x_n}(a;n) \qquad \forall h_n, a, n$$

$$\Pi \text{ είναι } S \iff \underbrace{\Pi_{h_n}(a;n) = \Pi_{x_n}(a;n)}_{\text{Μαρκοβιανή}} = \underbrace{\Pi_{x_n}(a)}_{\text{ανεξ. από το στάδιο}}$$

Μαρκοβιανή + ανεξ. από το στάδιο

2 Ως προς την τυχαιότητα

R (τυχαιοποιημένη) ⟶ ND

↓

D (προσδιοριστικές)

$$\Pi \text{ είναι } D \iff \Pi_{h_n}(a;n) = \begin{cases} 1, & \text{για } a = a^*(h_n;n) \\ 0, & \text{διαφ.} \end{cases}$$

Έχουμε τις κλάσεις πολιτικών:

$$\begin{tikzcd}
 & HR \arrow[dl] \arrow[dr] & & \\
HD & & MR \arrow[dl] \arrow[dr] & \\
 & \boxed{MD} & & SR \arrow[dr] \\
 & & & & SD
\end{tikzcd}$$

Βασικό αποτέλεσμα στο Δ.Π με πεπερ. χρονικό ορίζοντα: Ξβέλτιστη Πολιτική τύπου MD

Πιθανότητες Πραγματοποιήσεων ιστορίας

Αν $P_{xy}(a;n)$, $\Pi_{hn}(a;n)$, $\Pi^{(t)}_{x_t} = Pr[X_t = x_t]$, τότε

$Pr[H_n = h_n = x_t a_t x_{t-1} a_{t-1} \cdots x_{n+1} a_{n+1} x_n] = \Pi^{(t)}_{x_t} \, \Pi_{x_t}(a_t;t) P_{x_t x_{t-1}}(a_t;t) \cdot \Pi_{x_t a_t x_{t-1}}(a_{t-1};t-1)$

$\cdot P_{x_{t-1} x_{t-2}}(a_{t-1};t-1) \cdots \Pi_{x_t a_t \ldots x_{n+1}}(a_{n+1};n+1) P_{x_{n+1} x_n}(a_{n+1};n+1)$

Παρατηρήσεις!

(i) Αν π μαρκοβιανή, τότε η $\{X_t, X_{t-1}, \ldots, X_0\}$ είναι ΜΑΣΧ

Πράγματι, $Pr[X_{n-1} = x_{n-1} \mid X_t = x_t, X_{t-1} = x_{t-1}, \ldots, X_n = x_n] = \sum_{a_n \in \mathcal{A}(x_n,n)} \Pi_{x_n}(a_n;n) P_{x_n x_{n-1}}(a_n;n)$

(ii) Αν η π είναι στάσιμη, τότε η $\{X_t, X_{t-1}, \ldots, X_0\}$ και $P_{xy}(a;n)$ ανεξ. του n είναι ομογενής ΜΑΔΧ

Μάθημα 14ο (22/11/18)

Θεωρία Δ.Π σε πεπερασμένο χρονικό ορίζοντα

1. Βασικές έννοιες

$t = \#$ σταδίων

$\mathcal{X}$ = χώρος καταστάσεων

$\mathcal{A}$ = χώρος αποφάσεων

n = στάδιο $n = t, t-1, t-2, \ldots 0$

x_n = κατάσταση στο στάδιο n

a_n = απόφαση στο στάδιο n

$\mathcal{A}(x_n;n)$ = δεσμευμένη απόφαση της x_n στο στάδιο n

$P_{x_n x_{n-1}}(a_n;n)$ = πιθανότητα μετάβασης προς την x_{n-1} δεδομένων n, x_n, a_n

$c(x_n, a_n; n)$ = άμεσο κόστος λήψης a_n στην κατάσταση x_n στο στάδιο n

$c(x_0; 0)$ = τερματικό κόστος στην κατάσταση x

$h_n = \underbrace{x_t a_t x_{t-1} a_{t-1} \ldots x_{n+1} a_{n+1} x_n}_{\text{Πραγματοποίηση της ιστορίας μέχρι το στάδιο } n}$

π: πολιτική

$\pi = (\pi(t), \pi(t-1), \dots, \pi(1))$

$\pi(n)$ = Σύνολο ιστοριών στο στάδιο n ⟶ Σύνολο συναρτήσεων πιθ. στις αποφάσεις

$\pi_{h_n}(a_n; n)$ = πιθ. λήψης της a_n αν έχω παρατηρήσει h_n μέχρι το στάδιο n

h_n: $x_t a_t \dots x_{n+1} a_{n+1} x_n$

$(\pi_{h_n}(a_n;n) : a_n \in A(x_n;n))$ συνάρτηση πιθ.

2. Αξία πολιτικής

$$\underbrace{U_t^{\pi}(x_t) = E_\pi\left[\sum_{n=1}^{t} c(X_n, A_n; n) + c(X_0; 0) \mid X_t = x_t\right]}_{\text{μέσο συνολικό κόστος υπό την πολιτική } \pi \text{ ξεκινώντας από την κατάσταση } x_t}$$: συνάρτηση ολικής αξίας πολιτικής π

(π: πολιτική)

$n: t, t-1, \dots, 0$: $$\underbrace{U_n^{\pi}(h_n) = E_\pi\left[\sum_{k=1}^{n} c(X_k, A_k, k) + c(X_0; 0) \mid H_n = h_n\right]}_{\text{μέσο υπολειπόμενο κόστος από το στάδιο } n \text{ και μετά υπό την } \pi \text{ αν έχω παρατηρήσει } h_n}$$: συνάρτηση (μερικής) αξίας π

h_n: $x_t a_t \dots x_{n+1} a_{n+1} x_n$

__Πρόταση__

π πολιτική, τότε $U_0^{\pi}(h_0) = c(x_0; 0)$

$$U_n^{\pi}(h_n) = \sum_{a_n \in A(x_n;n)} \pi_{h_n}(a_n;n)\left(c(x_n, a_n; n) + \sum_{x_{n-1}} P_{x_n x_{n-1}}(a_n;n)\, U_{n-1}^{\pi}(h_n a_n x_{n-1})\right)$$

__Απόδειξη__

Θ.Ο.Π με δέσμευση στην A_n, X_{n-1}

__Συνάρτηση βέλτιστης τιμής__

$V_n(x_n)$: ελάχιστο υπολειπόμενο κόστος όταν στο στάδιο n παρατηρώ κατάσταση x_n $= \inf_{\pi, x_t, a_t, \dots, x_{n+1}, a_{n+1}} U_n^{\pi}(x_t a_t x_{t-1} a_{t-1} \dots x_{n+1} a_{n+1} x_n)$

3. Βασικό Αποτέλεσμα

Υπάρχει βέλτιστη πολιτική στην κλάση MD (Μαρκοβιανή, ντετερμινιστική)
δηλαδή $\exists \pi^*$: $V_n(x_n) = u_n^{\pi^*}(x_t a_t \dots x_{n+1} a_{n+1} x_n)$ $\forall x_t, a_t, \dots x_{n+1}, a_{n+1}, x_n$ και
$$\pi^*_{h_n}(a_n;n) = \pi^*_{x_n}(a_n;n) = \begin{cases} 1, & a_n = a_n^*(x_n) \\ 0, & \text{διαφ} \end{cases}$$

Λήμμα

Η $V_n(x_n)$ ικανοποιεί την ανισότητα:
$$V_n(x_n) \geq \min_{a_n \in \mathcal{A}(x_n;n)} \Big(\underbrace{c(x_n,a_n;n)}_{\text{άμεσο κόστος}} + \underbrace{\sum_{x_{n-1}} P_{x_n x_{n-1}}(a_n;n) V_{n-1}(x_{n-1})}_{\text{μελλοντικό κόστος}} \Big), \quad 1 \leq n \leq t, \ x_n \in \mathcal{X}$$

Απόδειξη

Παίρνω μια πολιτική π και $x_t, a_t, x_{t-1}, a_{t-1}, \dots, x_{n+1}, a_{n+1}$

Έστω $h_n = x_t a_t x_{t-1} a_{t-1} \dots x_{n+1} a_{n+1} x_n$
$$u_n^{\pi}(h_n) = \sum_{a_n} \pi_{h_n}(a_n;n) \Big(c(x_n,a_n;n) + \sum_{x_{n-1}} P_{x_n x_{n-1}}(a_n;n) u_{n-1}^{\pi}(h_n a_n x_{n-1}) \Big)$$
$$\overset{\text{ορσ } V_{n-1}}{\geq} \sum_{a_n} \pi_{h_n}(a_n;n) \Big(c(x_n,a_n;n) + \sum_{x_{n-1}} P_{x_n x_{n-1}}(a_n;a) V_{n-1}(x_{n-1}) \Big)$$
$$\geq \underbrace{\sum_{a_n} \pi_{h_n}(a_n;a)}_{1} \cdot \min_{a_n'} \Big(c(x_n,a_n';n) + \sum_{x_{n-1}} P_{x_n x_{n-1}}(a_n';a) V_{n-1}(x_{n-1}) \Big)$$
$$= \min_{a_n'} (- \ \| \ -)$$

Παίρνοντας inf ως προς $\pi, x_t, a_t \dots x_{n+1} a_{n+1}$ OK

Θεώρημα

Έστω $u_n(x_n)$ ακολουθία συναρτήσεων που ορίζεται αναδρομικά ως εξής:
$$u_0^*(x_0) = c(x_0;0)$$
$$u_n^*(x_n) = \min_{a_n \in \mathcal{A}(x_n;n)} \Big(c(x_n,a_n;n) + \sum_{x_{n-1}} P_{x_n x_{n-1}}(a_n;n) u_{n-1}^*(x_{n-1}) \Big), \quad 1 \leq n \leq t, \ x_n \in \mathcal{X}$$
$$= c(x_n, a_n^*(x_n);n) + \sum_{x_{n-1}} P_{x_n x_{n-1}}(a_n^*(x_n);n) u_{n-1}^*(x_{n-1})$$
Έστω $\pi^* = (\pi^*(t), \dots, \pi^*(1))$ η MD πολιτική με $\pi^*_{h_n}(a_n;n) = \begin{cases} 1, & \text{αν } a_n = a_n^*(x_n) \\ 0, & \text{διαφ} \end{cases}$

Τότε, $u_n^{\pi}(h_n) = u_n^*(x_n) = V_n(x_n)$

Άρα, π^* είναι βέλτιστη

Απόδειξη

Επαγωγή στο n

$n=0$ $\quad U_0^{\pi^*}(h_0) = U_0^*(x_0) = V_0(x_0) = c(x_0;0)$

εξ. ορισμού $\quad$ πρόταση $\quad$ άμεσο

Έστω ότι ισχύει για $n-1$ $\quad U_{n-1}^{\pi^*}(h_{n-1}) = U_{n-1}^*(x_{n-1}) = V_{n-1}(x_{n-1})$

Έχω $h_n = x_t a_t \dots x_{n+1} a_{n+1} x_n$

$$U_n^{\pi^*}(h_n) = \sum_{a_n} \pi^*_{h_n}(a_n;a)\Big(c(x_n,a_n;n) + \sum_{x_{n-1}} P_{x_n x_{n-1}}(a_n;n)\underbrace{U_{n-1}^{\pi^*}(h_n a_n x_{n-1})}_{\substack{\parallel \text{ επαγ. υποθ.} \\ U^*_{n-1}(x_{n-1}) \\ \parallel \\ V_{n-1}(x_{n-1})}}\Big)$$

Άρα, $U_n^{\pi^*}(h_n) = c(x_n, a_n^*(x_n);n) + \sum_{x_{n-1}} P_{x_n x_{n-1}}(a_n^*(x_n);n)\underbrace{U_{n-1}^{\pi^*}(h_n a_n^*(x_n) x_{n-1})}_{\substack{\parallel \\ U^*_{n-1}(x_{n-1}) = V_{n-1}(x_{n-1})}}$

Άρα, $U_n^{\pi^*}(h_n) = U_n^*(x_n)$ (λόγω της 2ης ισότητας του ορ. της $U_n^*(x_n)$)

Επίσης, $V_n(x_n) = \inf_{\pi, x_t \dots x_{n+1}, a_{n+1}} U_n^{\pi}(x_t a_t \dots x_{n+1} a_{n+1} x_n) \le U^{\pi^*}(h_n)$

$$= c(x_n, a_n^*(x_n);n) + \sum_{x_{n-1}} P_{x_n x_{n-1}}(a_n^*(x_n);n) V_{n-1}(x_{n-1})$$

$$= \min\Big(c(x_n,a_n;n) + \sum_{x_{n-1}} P_{x_n x_{n-1}}(a_n;n) V_{n-1}(x_{n-1})\Big)$$

Από λήμμα & τελευταία ισότητα $\Rightarrow V_n(x_n) = \min\Big(c(x_n,a_n;n) + \sum_{x_{n-1}} P_{x_n x_{n-1}}(a_n;n) V_{n-1}(x_{n-1})\Big)$

Άρα, $U_n^*(x_n) = V_n(x_n)$