Ενότητα 6 - Δυναμικός Προγραμματισμός

15-12-2023

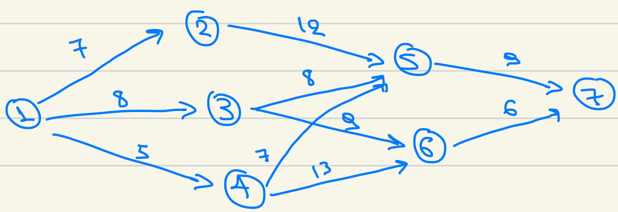
Ασχολούμαστε με προβλήματα που αποτελούνται από στάδια και σε κάθε στάδιο παίρνουμε μία απόφαση.

Ιδέα: Σπάμε το πρόβλημα σε υποπροβλήματα (στάδια) και κάθε ένα στάδιο έχει μία μεταβλητή απόφασης.

Παράδειγμα 1 (Πρόβλημα ελάχιστης διαδρομής)

Θέλουμε την ελάχιστη διαδρομή από την πόλη 1 στην πόλη 7.

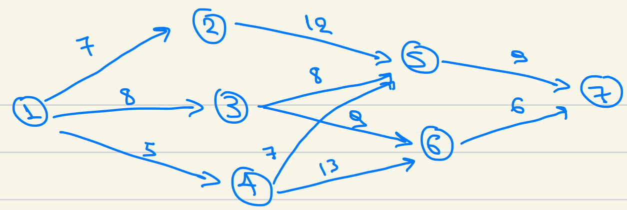
Οι διαδρομές περνάν από τις πόλεις 2 - 6.



Λύση:

Παρατηρήσεις:

- Κάθε διαδρομή αποτελείται από 3 βήματα, άρα θα πάρουμε 3 διαδοχικές αποφάσεις.
- Σε κάθε βήμα, ανάλογα με την πόλη (δηλαδή, με την κατάσταση) που βρίσκομαι, μπορώ να πάρω κάποιες αποφάσεις.
- Κάθε απόφαση έχει κάποιο κόστος (άμεσο) και με οδηγεί σε μία επόμενη κατάσταση.



Στ.1 Στ.2 Στ.3

$v(x)$ = μήκος ελάχιστης διαδρομής από την πόλη x στην 7.

Θέλω το $v(1)$

Στάδιο 3

$v(5) = 9$, $a^*(5) = 7$

η βέλτιστη απόφαση όταν είμαι στην πόλη (κατάσταση) 5 είναι να πάω στην 7.

$v(6) = 6$, $a^*(6) = 7$

Στάδιο 2

$v(2) = 12 + v(5) = 12+9=21$, $a^*(2) = 5$

$v(3) = \min\{\underbrace{8+v(5)}_{\text{αποφ } 5}, \underbrace{9+v(6)}_{\text{αποφ } 6}\} =$

$\min\{8+9, 9+6\} = \min\{17, 15\} = 15$, $a^*(3) = 6$

$v(4) = \min\{\underbrace{7+v(5)}_{\text{αποφ } 5}, \underbrace{13+v(6)}_{\text{αποφ } 6}\} =$

$= \min\{7+9, 13+6\} =$

$\min\{16, 19\} = 16$, $a^*(4) = 5$

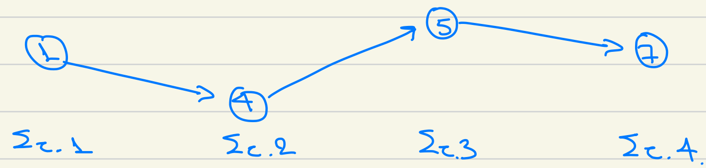
Στάδιο 1

$v(1) = \min\{\underbrace{7+v(2)}_{\text{αποφ } 2}, \underbrace{8+v(3)}_{\text{αποφ } 3}, \underbrace{5+v(4)}_{\text{αποφ } 4}\} =$

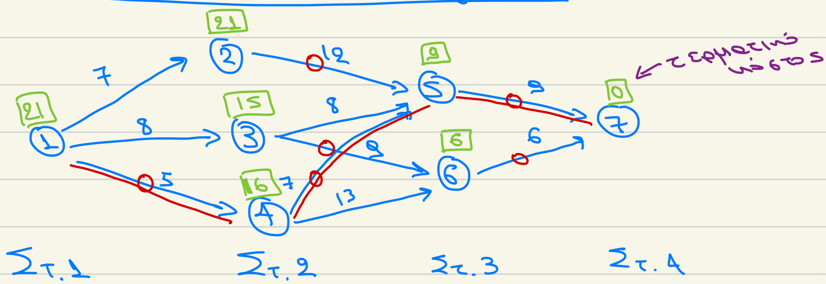
$= \min\{7+21, 8+15, 5+16\} =$

$\min\{28, 23, 21\} = 21$, $a^*(1) = 4$.

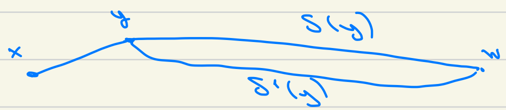
Βρήκαμε $u(1) = 21$



Επίλυση πάνω σε διάγραμμα:



Αρχή βελτιστοποίησης Bellman



Αν $\delta(x) = ((x,y), \delta(y))$ είναι βέλτιστη από το x στο w, τότε και η $\delta(y)$ θα είναι βέλτιστη από το y στο w.

Απόδειξη

Έστω ότι $\delta(y)$ δεν είναι βέλτιστη από το y στο w και υπάρχει $\delta'(y)$ με μικρότερο κόστος. Τότε $\delta'(x) = ((x,y), \delta'(y))$ θα είχε μικρότερο κόστος από τη $\delta(x)$. Άτοπο.

Συμβολισμοί - Ορισμοί

Σε ένα πρόβλημα δυναμικού προγραμματισμού πεπερασμένου ορίζοντα με ντετερμινιστική δυναμική ορίζουμε

- Στάδια: $t = 1, 2, \ldots, N, N+1$

  N = μήκος ορίζοντα (# διαδοχικών αποφάσεων)

  $N+1$: τερματικό στάδιο

- Καταστάσεις: x_t : κατάσταση στο στάδιο t

  S_t = χώρος καταστάσεων στο στάδιο t

  (Θα ασχοληθούμε με π.δ.π. με πεπερασμένο χώρο καταστάσεων).

- Αποφάσεις: a_t : απόφαση στο στάδιο t

  $D_t(x_t)$ = δέσμη δυνατών αποφάσεων στο στάδιο t όταν βρίσκομαι στην x_t.

  (Θα ασχοληθούμε με π.δ.π με πεπερασμένη δέσμη δυνατών αποφάσεων)

- Δυναμική συστήματος: Ο μηχανισμός που ορίζει τη μεταβολή της κατάστασης σαν αποτέλεσμα της απόφασης

$$x_{t+1} = g_t(x_t, a_t)$$

- Άμεσο κέρδος/κόστος: $c_t(x_t, a_t)$

- Τερματικό κέρδος/κόστος: $\hat{c}(x_{N+1})$

Εξισώσεις βελτιστοποίησης

$$v(t, x_t) = \min_{a_t \in D_t(x_t)}/\max \left\{ C_t(x_t, a_t) + v(t+1, g_t(x_t, a_t)) \right\}$$

$t = 1, 2, \ldots, N$

τερματικές συνθήκες

$$v(N+1, x_{N+1}) = \hat{c}(x_{N+1})$$

Παράδειγμα 2

Ένας φοιτητής πρέπει να επιλέξει 7 μαθήματα από 4 κατηγορίες, με τουλάχιστον ένα μάθημα από κάθε κατηγορία. Ο παρακάτω πίνακας δίνει ένα σκορ που μετράει τη "γνώση" που θα αποκομίσει παίρνοντας κάποιον αριθμό μαθημάτων από κάθε κατηγορία.

	# μαθημάτων			
Κατηγορία (t)	1	2	3	4
1	25	50	60	80
2	20	70	90	100
3	40	60	80	100
4	10	20	30	40

Πόσα μαθήματα πρέπει να επιλέξει από κάθε κατηγορία για να μεγιστοποιήσει το σκορ;

Λύση:

Μοντελοποίηση ως π.δ.π:

- **Στάδια:** $t = 1, 2, 3, 4, 5$ ← τερματικό

  $N = 4$

  Στο στάδιο t αποφασίζω πόσα μαθήματα θα πάρει από την κατηγορία t.

- <u>Κατάσταση</u>: x_t = # μαθημάτων που απομένουν στην αρχή του σταδίου t.

- <u>Αποφάσεις</u>: a_t = # μαθημάτων που θα πάρω από την κατηγορία t (στο στάδιο t)

$D_t(x_t)$ $\qquad a_t \ge 1$

$\qquad a_t \le x_t - (4-t)$

$$D_t(x_t) = \{a_t \in \mathbb{N} : 1 \le a_t \le x_t - (4-t)\}$$

- <u>Δυναμική συστήματος</u>: $x_{t+1} = g_t(x_t, a_t) \Rightarrow$

$x_{t+1} = x_t - a_t$

- <u>Άμεσο κέρδος</u>: $c_t(x_t, a_t) = R(t, a_t)$ = όφελος αν πάρω a_t μαθήματα από την κατηγορία t.
(Δίνεται από τον πίνακα)

- <u>Τερματικό κέρδος</u>: $\hat{c}(x_5) = 0$

- <u>Εξισώσεις βελτιστοποίησης</u>

$v(t, x_t)$ = μέγιστο κέρδος από το στάδιο t μέχρι το τέλος του χρονικού ορίζοντα, όταν στην αρχή του σταδίου t βρίσκομαι στην κατάσταση x_t.

$$v(t, x_t) = \max_{a_t \in D_t(x_t)} \left\{ c_t(x_t, a_t) + v(t+1, \underbrace{g_t(x_t, a_t)}_{x_{t+1}}) \right\},$$

$t = 1, 2, \dots, 4$

Τερματικές συνθήκες

$$v(5, x_5) = \hat{c}(x_5) = 0$$

Επίλυση σε διάγραμμα: