Μάθημα 5

0 Εισαγωγικά

Εκτός από την απλή περιγραφή ενός συστήματος, η εξέλιξη του οποίου περιγράφεται από μία μ.α., πολλές φορές μπορούμε να παρέμβουμε στο σύστημα σε κάθε χρονική περίοδο και να παίρνουμε αποφάσεις που αλλάζουν τις πιθανότητες μετάβασης της μ.α. και τις τιμές της συνάρτησης κόστους που αντιστοιχεί στο συγκεκριμένο πρόβλημα.

Για κάθε μία από τις καταστάσεις, πολλές φορές έχουμε ένα σύνολο εναλλακτικών ενεργειών που θα μπορούσαμε να αποφασίσουμε να πράξαμε και ο στόχος είναι να κινηθούμε βέλτιστα, δηλ. να βρούμε τη βέλτιστη διαδικασία αποφάσεων που συνδέεται με την εκάστοτε μ.α. και συνάρτηση κόστους. Μία τέτοια διαδικασία αποφάσεων λέγεται μαρκοβιανή διαδικασία αποφάσεων.

Μαρκοβιανές Διαδικασίες Αποφάσεων (μ.δ.α.)

1 Παράδειγμα

Ένας κατασκευαστής έχει μία κύρια μηχανή στην καρδιά μιας παραγωγικής διαδικασίας. Στο τέλος κάθε εβδομάδας γίνεται επιθεώρηση και η μηχανή βρίσκεται σε κάποια από αυτές τις καταστάσεις:

0 ⟶ λειτουργεί σαν καινούρια

1 ⟶ λειτουργική με μικρή χειροτέρευση

2 ⟶ λειτουργική με μεγάλη χειροτέρευση

3 ⟶ όχι λειτουργική

[βλάβη ή μη αποδεκτή ποιότητα παραγωγής]

Χωρίς κάποια παρέμβαση έχει διαπιστωθεί εμπειρικά ότι η κατάσταση της μηχ. εξελίσσεται ως μ.α. με πίνακα μετάβασης

$$P = \begin{pmatrix} 0 & \frac{7}{8} & \frac{1}{16} & \frac{1}{16} \\ 0 & \frac{3}{4} & \frac{1}{8} & \frac{1}{8} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Παρατηρείστε ότι η κατάσταση 3 είναι απορροφητική.
Εξετάζεται ένα σύνολο στρατηγικών και υπολογίζονται <u>αναμενόμενα κόστη</u>

<u>Πίνακας 1</u>

/εβδομάδα

απόφαση	καταστ.	Αναμενόμενο κόστος λόγω παραγ. ελαττ. αντικ.	Κόστος συντήρησης	Κόστος Χαμένης Παραγωγής	Συνολικό κόστος/εβδομάδα
1: ΔΕΝ ΚΑΝΩ ΤΙΠΟΤΑ	0	0	0	0	0
	1	1,000	0	0	1.000
	2	3,000	0	0	3.000
2: ΕΠΙΣΚΕΥΗ	2	0	2.000	2.000	4.000
3: ΑΝΤΙΚΑΤΑΣΤΑΣΗ	1,2,3	0	4.000	2.000	6.000

② Μοντέλο μ.δ.α.

1) Παρατηρούμε την κατάσταση i μετά από κάθε μετάβαση.

2) Μετά την παρατήρηση, μία απόφαση (ενέργεια) k επιλέγεται μέσα από ένα σύνολο K δυνατών αποφάσεων $(k = 1, 2, \dots, K)$. Κάποιες μπορεί να αποκλείονται.

3) Αν η απόφαση $d_i = k$ [$d_i \to$ decision for the i-th state], τότε αυτό επιφέρει ένα άμεσο κόστος με μέση τιμή C_{ik}.

4) Η απόφαση $d_i = k$ καθορίζει επίσης τις πιθανότητες μετάβασης του επόμενου βήματος. Συμβολικά γράφουμε $P_{ij}(k)$ [αφού τα P_{ij} εξαρτώνται από το k]

5) Κάθε προσδιορισμός αποφάσεων $(d_0, d_1, \dots, d_M)$ καθορίζει μία <u>πολιτική R</u> για τη μ.δ.α.
[Μία πολιτική λοιπόν είναι ένας συνδυασμός αποφάσεων, μία για κάθε κατάσταση]

6) Ο στόχος είναι η εύρεση <u>βέλτιστης πολιτικής</u> σύμφωνα με κάποιο κριτήριο κόστους. Εμείς θα χρησιμοποιήσουμε το μέσο κόστος ανά χρονική μονάδα σε στασιμότητα.

<u>Παρατήρηση</u>

Τα είδη της πολιτικής που χρησιμοποιούμε είναι:

<u>στάσιμη</u> πολιτική : $d_i(t) = k_i$ (δηλ. ανεξάρτητη της χρονικής στιγμής t)

<u>ντετερμινιστική</u> πολιτική : $i \xrightarrow{\text{decision}} k_i$ (ντετερμινιστικά, χωρίς κατανομή πιθανότητας)

Συμβολικά γράφουμε και $d_i(R)$, όταν θέλουμε να δηλώσουμε την απόφαση που παίρνουμε στην κατάσταση i κάτω από πολιτική R.

Είναι σημαντικό λοιπόν να κατανοηθεί ότι <u>συγκρίνονται οι πολιτικές</u> και όχι οι αποφάσεις μεμονωμένα.

Πίνακας 2

πολιτική	Περιγραφή	$d_0(R)$	$d_1(R)$	$d_2(R)$	$d_3(R)$
R_a	Αντικατάσταση στην κατ. 3	1	1	1	3
R_b	Αντικατ. στην 3 και Επισκευή στη 2	1	1	2	3
R_c	Αντικατ. στις 2 και 3	1	1	3	3
R_d	Αντικατ. στις 1,2 και 3	1	3	3	3

Συμπεραίνουμε ότι

$$\left(d_0(R_a),\ d_1(R_a),\ d_2(R_a),\ d_3(R_a)\right) = (1, 1, 1, 3)$$

$$\left(d_0(R_b),\ d_1(R_b),\ d_2(R_b),\ d_3(R_b)\right) = (1, 1, 2, 3)$$

κ.τ.λ.

Η <u>εξαντλητική απαρίθμηση</u> αναφέρεται στην καταγραφή όλων των δυνατών πολιτικών και υπολογισμός της μέσης τιμής $E(C) = \sum_{i=0}^{M} C_{ik}\pi_i$, όπου $k = d_i(R)$, $\forall R$.

[απλοποίηση της έκφρασης

$$E_{\pi(R)}(C) = \sum_{i=0}^{M} C_{i,d_i(R)}\,\pi_i(R)$$],

αφού το κόστος και η στάσιμη κατανομή εξαρτώνται από την πολιτική.

$\boxed{\text{Επιλέγουμε το } R \text{ που ελαχιστοποιεί το } E(C)}$

Η μέθοδος αυτή προϋποθέτει τον υπολογισμό της στάσιμης κατανομής $\forall$ πολιτική R.

<u>Παράδειγμα</u>

Τροποποιούμε την αρχική μ.α. ανάλογα με τις αποφάσεις που παίρνουμε για κάθε πολιτική και προκύπτουν

R_a: $P = \begin{pmatrix} 0 & 7/8 & 1/16 & 1/16 \\ 0 & 3/4 & 1/8 & 1/8 \\ 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \end{pmatrix}$ R_b: $P = \begin{pmatrix} 0 & 7/8 & 1/16 & 1/16 \\ 0 & 3/4 & 1/8 & 1/8 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$

R_c: $P = \begin{pmatrix} 0 & 7/8 & 1/16 & 1/16 \\ 0 & 3/4 & 1/8 & 1/8 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$ R_d: $P = \begin{pmatrix} 0 & 7/8 & 1/16 & 1/16 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$

πολιτική	στάσιμο διάνυσμα	$E(C)$ σε χιλιάδες
R_a	$\left(\frac{2}{13}, \frac{7}{13}, \frac{2}{13}, \frac{2}{13}\right)$	$\frac{1}{13}\left(2\times 0 + 7\times 1 + 2\times 3 + 2\times 6\right) = 1,923$
R_b	$\left(\frac{2}{21}, \frac{5}{7}, \frac{2}{21}, \frac{2}{21}\right)$	$\ldots = 1,667$
R_c	$\left(\frac{2}{11}, \frac{7}{11}, \frac{1}{11}, \frac{1}{11}\right)$	$\ldots = 1,727$
R_d	$\left(\frac{1}{2}, \frac{7}{16}, \frac{1}{32}, \frac{1}{32}\right)$	$\ldots = 3,000$

<u>Παρατ</u>: Τα C_{ik} προκύπτουν από την τελευταία στήλη του Πιν. 1

Η βέλτιστη πολιτική είναι η R_b αφού ελαχιστοποιείται το $E(C)$.

4 Μέθοδοι Επίλυσης – Αλγόριθμος Βελτίωσης Πολιτικής

Μπορούμε ν.δ.ο. $\forall$ πολιτική R, $\exists\, g(R), U_i(R)$, $0 \le i \le M$:

$$\boxed{g(R) + U_i(R) = C_{ik} + \sum_{j=0}^{M} P_{ij}(k)\, U_j(R), \quad \forall i = 0, 1, \ldots, M}$$

- Λόγω σύμβασης θέτουμε $U_M(R) = 0$ και έτσι έχουμε $M+1$ εξισώσεις με $M+1$ αγνώστους $\left(g(R), U_0(R), U_1(R), \ldots, U_{M-1}(R)\right)$
- Δείτε σελ. 1064–65 για μία εμπειρική αιτιολόγηση και ερμηνεία των ποσοτήτων:

  $g(R) = E_{\pi(R)}(C)$ και $U_i(R)$: η "επίδραση" στο συνολικό αναμενόμενο κόστος, λόγω εκκίνησης στην κατάσταση i.
- Οι εξισώσεις αυτές είναι στη βάση κατασκευής του επόμενου αλγορίθμου.

Αλγόριθμος Βελτίωσης Πολιτικής (Α.Β.Π).

<u>Αρχικοποίηση</u> $(n=1)$: Επιλέγουμε αυθαίρετα μία αρχική πολιτική R_1.

<u>Επανάληψη n</u> : Συνίσταται από 2 βήματα

<u>Βήμα 1</u> : <u>Βήμα Προσδιορισμού Τιμών</u> (Β.Π.Τ.).

Για την πολιτική R_n, χρησιμοποιούμε τα $P_{ij}(k)$, C_{ik} και $U_M(R_n)=0$, για να λύσουμε το σύστημα των εξισώσεων :

$$g(R_n) = C_{ik} + \sum_{j=0}^{M} P_{ij}(k)\, U_j(R_n) - U_i(R_n), \ \forall i=0,1,\dots M$$

με $M+1$ αγνώστους τα $g(R_n), U_0(R_n), U_1(R_n), \dots, U_{M-1}(R_n)$.

<u>Βήμα 2</u> : <u>Βήμα Βελτίωσης Πολιτικής</u> (Β.Β.Π).

$\forall$ κατάσταση $i=0,1,\dots,M$,

<u>ελαχιστοποιούμε</u> την ποσότητα

$$C_{ik} + \sum_{j=0}^{M} P_{ij}(k)\, U_j(R_n) - U_i(R_n),$$

ως προς τα εφικτά $k \in \{1, 2, \dots, K\}$,

βρίσκοντας έτσι τη βέλτιστη απόφαση k_i^*, $\forall$ καταστ. i, και θέτουμε $d_i(R_{n+1}) = k_i^*$.

Με αυτόν τον τρόπο καθορίζεται μία καλύτερη πολιτική R_{n+1}.

Έλεγχος Βελτιστότητας

Η παρούσα πολιτική R_{n+1} είναι βέλτιστη, αν συμπίπτει με την R_n, και τότε σταματάει ο αλγόριθμος. Διαφορετικά, επαναλαμβάνουμε τη διαδικασία κάνοντας μία ακόμα επανάληψη.

Ιδιότητες του Αλγορίθμου

1) $g(R_{n+1}) \leq g(R_n)$, $\forall n = 1, 2, \ldots$

2) Ο αλγόριθμος τερματίζει με μία βέλτιστη πολιτική σε πεπερασμένο πλήθος επαναλήψεων.

Παράδειγμα – Επίλυση με Α.Β.Π.

Αρχικοποίηση : Θέτουμε $R_1 = R_a$ (αντικατάσταση στη 3 και τίποτα στις άλλες)

Υπενθύμιση

Πολιτική $R_1 = R_a$

Κατάστ.	Απόφ.
0	1
1	1
2	1
3	3

Πίνακας μετάβασης για R_1

$$\begin{pmatrix} 0 & 7/8 & 1/16 & 1/16 \\ 0 & 3/4 & 1/8 & 1/8 \\ 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

κόστοι $C_{ik}(R_1)$

Κατάστ.	C_{ik}
0	0
1	1
2	3
3	6

Επανάληψη 1

• Β.Π.Τ. $[\forall i : g = C_{ik} + P_{i0}(k)U_0 + P_{i1}U_1 + \ldots + P_{i3}(k)\cancel{U_3}^{0} - U_i]$

$$\left.\begin{array}{ll} i=0: & g = 0 + 0 + \frac{7}{8}U_1 + \frac{1}{16}U_2 - U_0 \\ i=1: & g = 1 + 0 + \frac{3}{4}U_1 + \frac{1}{8}U_2 - U_1 \\ i=2: & g = 3 + 0 + 0 + \frac{1}{2}U_2 - U_2 \\ i=3: & g = 6 + 1 \cdot U_0 \end{array}\right\} \Rightarrow$$

Η λύση που παίρνουμε είναι :

$\underset{g(R_1)}{\underset{\downarrow}{g}} = 1,923$, $\underset{U_0(R_1)}{\underset{\downarrow}{U_0}} = -4,077$, $\underset{U_1(R_1)}{\underset{\downarrow}{U_1}} = -2,615$, $\underset{U_2(R_1)}{\underset{\downarrow}{U_2}} = 2,154$

• <u>Β.Β.Π.</u> (ελαχιστοποίηση ως προς k)

$i=0$: $C_{0k} - P_{00}(k)\times 4,077 - P_{01}(k)\times 2,615 + P_{02}(k)\times 2,154 + 4,077$ (0)

$i=1$: $C_{1k} - P_{10}(k)\times 4,077 - P_{11}(k)\times 2,615 + P_{12}(k)\times 2,154 + 2,615$ (1)

$i=2$: $C_{2k} - P_{20}(k)\times 4,077 - P_{21}(k)\times 2,615 + P_{22}(k)\times 2,154 - 2,154$ (2)

$i=3$: $C_{3k} - P_{30}(k)\times 4,077 - P_{31}(k)\times 2,615 + P_{32}(k)\times 2,154$

Απο τον Πίνακα 2 βλέπουμε τις επιλογές για το $d_i(R)$, $\forall i$

• $\boxed{i=0}$

μόνη επιλογή $d_0(R) = 1 \Rightarrow d_0(R_2) = 1$.

• $\boxed{i=1}$

$d_1(R) \in \{1,3\}$

απόφαση k	C_{1k}	$P_{10}(k)$	$P_{11}(k)$	$P_{12}(k)$	τιμή (σχέση (1))
1	1	0	3/4	1/8	1,923 ← Ελάχιστο
3	6	1	0	0	4,538

$\Rightarrow d_1(R_2) = 1$

• $\boxed{i=2}$

$d_2(R) \in \{1, 2, 3\}$

απόφαση κ	C_{2k}	$P_{20}(k)$	$P_{21}(k)$	$P_{22}(k)$	τιμή (σχέση (2))
1	4	0	0	1/2	1,923
2	3	0	1	0	− 0,769
3	6	1	0	0	− 0,231

$\Rightarrow \quad d_2(R_2) = 2$

• $\boxed{i=3}$

μοναδική επιλογή $d_3(R) = 3 \Rightarrow d_3(R_2) = 3$

έλεγχος βελτιστότητας $R_1 = R_a \downarrow (1, 1, 1, 3)$, ενώ $R_2 = R_b \downarrow (1, 1, 2, 3)$ $\Big\} \Rightarrow$

$R_2 \neq R_1$ και άρα συνεχίζουμε.

Επανάληψη 2 : Θέτουμε $R_2 = R_b$

• Β.Π.Τ.

πολιτική / πίνακας μετάβασης / κόστοι C_{ik} (R_2)

+ σύστημα εξισώσεων $\Rightarrow g(R_2) = 1,667$

$v_0(R_2) = -4,333$, $v_1(R_2) = -3$, $v_2(R_2) = -0,667$

• Β.Β.Π $\Rightarrow R_3 = R_2 \Rightarrow R^* = R_b$, με

ελάχιστο μέσο κόστος/χρονική περίοδο σε σταθερή κατάσταση $\boxed{g^* = 1,667}$

+ Ασκ. 21.2.1.