Aalto University
School of Science

Partially observable Markov decision process (POMPD)

Jussi Leppinen
Presentation *15*
6.11.2020

MS-E2191 Graduate Seminar on Operations Research
Fall 2020

Contents

POMDP

- **First used by Drake (1962)**
  - Study of monitoring of a discrete Markov source through a memoryless noisy channel

- **Why partial information?**
  - Structural systems are not identical to their designs
  - Direct, accurate measurements are not possible

POMDP

- Describes the problem when the transition laws of the core process under different control actions are known but the true state of the core process cannot be identified with certainty
  - Observable states Y (state space M) and true states X (state space S)
  - Control action during time t: a_t

- Relationship matrix: $\boldsymbol{R}(a_{t-1}) = \left[r_{ij}(a_{t-1})\right] \in \mathbb{R}^{|M|\times|S|}$ where $r_{ij}(a_{t-1}) = P[Y_t = j|X_t = i, a_{t-1}]$
- Transition matrix during $[t, t+1)$ under control a_t: $\boldsymbol{P}(a_t) = \left[p_{ij}(a_t)\right] \in \mathbb{R}^{|S|\times|S|}$where $p_{ij}(a_t) = P[X_{t+1} = j|X_t = i, a_t]$

Information vector

- Information vector $\boldsymbol{q}(t) = \left(q_1(t), \dots, q_N(t)\right)$ where $q_i(0) = P[X_0 = i | d_t]$ where d_t is the history from 0 to t

- Information is updated: $\boldsymbol{q}_j^k(t+1) = (q_{j1}^k(t+1), \dots, q_{jN}^k(t+1))$ where

$$q_{ji}^k(t+1) = \frac{\sum_{l \in S} q_l(t) p_{li}(k) r_{ij}(k)}{\sum_{i \in S} [\sum_{l \in S} q_l(t) p_{li}(k)] r_{ij}(k)} := \frac{q(t) P_r^{j,k}}{\gamma(j | d_t, k)}$$

POMPD into MDP

- The sequence of information vectors $\{\boldsymbol{q}(t), t \in T\}$ is a Markov process for any fixed sequence of actions and it is completely "observable".
- Terminal cost $\boldsymbol{C}^0 = (C_1^0, \dots, C_N^0)^T$
- Immediate cost $\boldsymbol{C}_k = (C_{1k}, \dots, C_{Nk})^T$

- Value function for the problem

$$V^n\big(\boldsymbol{q}(t)\big) = \min_{k \in A}\left[\boldsymbol{q}(t)\boldsymbol{C_k} + \sum_{j \in M} \gamma(j|d_t, k)\, V^{n-1}\left(\boldsymbol{q}_j^k(t+1)\right)\right]$$

- For $n = 0$, $V^0\big(\boldsymbol{q}(t)\big) = \boldsymbol{q}(t)\boldsymbol{C}^0$

Multiple inspection strategies

- Inspection method is chosen from I
- Other costs (model specific)
  - Penalty cost $\boldsymbol{C}^P = (C_1^P, \dots, C_N^P)^T$
  - Inspection strategy cost $\boldsymbol{C}_l^I = \left(C_{l1}^I, \dots, C_{lN}^I\right)^T$
  - Discount factor β

- Recursive formulation

$$V^n\big(\boldsymbol{q}(t)\big) = \min_{k\in A, l\in I}\left[\boldsymbol{q}(t)\boldsymbol{C_k} + \beta\boldsymbol{q}(t)\boldsymbol{P}(k)(\boldsymbol{C}^P + \boldsymbol{C}_l^I) + \beta\sum_{j\in M(l)}\gamma(j|d_t, k, l)\, V^{n-1}\left(\boldsymbol{q}_j^{k,l}(t+1)\right)\right]$$

Example: Bridge repair

$$\mathbf{P}(1) = \begin{bmatrix} 0.80 & 0.13 & 0.02 & 0.00 & 0.05 \\ 0.00 & 0.70 & 0.17 & 0.05 & 0.08 \\ 0.00 & 0.00 & 0.75 & 0.15 & 0.10 \\ 0.00 & 0.00 & 0.00 & 0.60 & 0.40 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{bmatrix} \quad \mathbf{R}(1) = (1,1,1,1,1)^T$$

$$\mathbf{P}(2) = \begin{bmatrix} 0.80 & 0.13 & 0.02 & 0.00 & 0.05 \\ 0.00 & 0.80 & 0.10 & 0.02 & 0.08 \\ 0.00 & 0.00 & 0.80 & 0.10 & 0.10 \\ 0.00 & 0.00 & 0.00 & 0.60 & 0.40 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{bmatrix} \quad \mathbf{R}(2) = \begin{bmatrix} 0.80 & 0.20 & 0.00 \\ 0.20 & 0.60 & 0.20 \\ 0.05 & 0.70 & 0.25 \\ 0.00 & 0.30 & 0.70 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$$

$$\mathbf{P}(3) = \begin{bmatrix} 0.80 & 0.13 & 0.02 & 0.00 & 0.05 \\ 0.19 & 0.65 & 0.08 & 0.02 & 0.06 \\ 0.10 & 0.20 & 0.56 & 0.08 & 0.06 \\ 0.00 & 0.10 & 0.25 & 0.55 & 0.10 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{bmatrix} \quad \mathbf{R}(3) = \begin{bmatrix} 0.90 & 0.10 & 0.00 & 0.00 & 0.00 \\ 0.05 & 0.90 & 0.05 & 0.00 & 0.00 \\ 0.00 & 0.05 & 0.90 & 0.05 & 0.00 \\ 0.00 & 0.00 & 0.05 & 0.95 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{bmatrix}$$

$$\mathbf{P}(4) = \begin{bmatrix} 0.80 & 0.13 & 0.02 & 0.00 & 0.05 \\ 0.80 & 0.13 & 0.02 & 0.00 & 0.05 \\ 0.80 & 0.13 & 0.02 & 0.00 & 0.05 \\ 0.80 & 0.13 & 0.02 & 0.00 & 0.05 \\ 0.80 & 0.13 & 0.02 & 0.00 & 0.05 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} 0 & 5 & 25 & 40 \\ 0 & 8 & 80 & 120 \\ 0 & 15 & 100 & 550 \\ 300 & 320 & 450 & 800 \\ 2000 & 2050 & 2500 & 4000 \end{bmatrix}$$

$$\boldsymbol{C}_1^I = (0,0,0,0,0)^T$$
$$\boldsymbol{C}_2^I = (4,4,4,4,4)^T$$
$$\boldsymbol{C}_3^I = (18,18,18,18,18)^T$$
$$\boldsymbol{C}^P = (0,0,0,0,0)^T$$
$$\boldsymbol{C}^0 = (0,0,0,0,0)^T$$

α-vectors

$$V^n(\boldsymbol{q}(t)) = \min_{k\in A, l\in I}\left[\boldsymbol{q}(t)\boldsymbol{C}_k + \beta\boldsymbol{q}(t)\boldsymbol{P}(k)\boldsymbol{C}_l^I + \beta\sum_{j\in M(l)}\gamma(j|d_t,k,l)\,V^{n-1}\left(\boldsymbol{q}_j^{k,l}(t+1)\right)\right]$$

- Now $V^0(\boldsymbol{q}(t)) = \boldsymbol{q}(t)\boldsymbol{C}^0 = 0$
- Then V^1 can be calculated separately for each k and l
- For example when $k = 1$ and $l = 1$
  - $V^1(\boldsymbol{q}(t)) = \boldsymbol{q}(t)\boldsymbol{C}_1 + \beta\boldsymbol{q}(t)\boldsymbol{P}(1)\boldsymbol{C}_1^I = \boldsymbol{q}(t)(\boldsymbol{C}_1+\beta\boldsymbol{P}(1)\boldsymbol{C}_1^I) \coloneqq \boldsymbol{q}(t)\boldsymbol{\alpha}_1^{1,1}$
- Recursively, when $k = 2$ and $l = 1$
  - $V^2(\boldsymbol{q}(t)) = \boldsymbol{q}(t)\boldsymbol{C}_2 + \beta\boldsymbol{q}(t)\boldsymbol{P}(2)\boldsymbol{C}_1^I + \beta\sum_{j\in M(1)}\gamma(j|d_t,2,1)\,V^1\left(\frac{\boldsymbol{q}(t)\boldsymbol{P}_r^{j,2,1}}{\gamma(j|d_t,2,1)}\right) = \boldsymbol{q}(t)\boldsymbol{C}_2 + \beta\boldsymbol{q}(t)\boldsymbol{P}(2)\boldsymbol{C}_1^I + \beta\sum_{j\in M(1)}\boldsymbol{q}(t)\boldsymbol{P}_r^{j,2,1}\ \boldsymbol{\alpha}_1^{2,1} = \boldsymbol{q}(t)(\boldsymbol{C}_2+\beta\boldsymbol{P}(2)\boldsymbol{C}_1^I + \beta\sum_{j\in M(1)}\boldsymbol{P}_r^{j,2,1}\ \boldsymbol{\alpha}_1^{2,1}) \coloneqq \boldsymbol{q}(t)\boldsymbol{\alpha}_2^{2,1}$
- General recursive formulation is found in White (1993)

Example: Optimal decisions

- Calculate the α-vectors recursively for each action and inspection possibilities for all the needed time steps (n)
- When the information $\boldsymbol{q}(t)$ is known, choose the optimal maintenance decision k' and inspection strategy l' from the corresponding α-vector giving smallest value of $\boldsymbol{q}(t)\boldsymbol{\alpha}_n^{k,l}$
- Update the information according to the outcome of the inspection j.
- Choose optimal decisions based on the value of $\boldsymbol{q}_j^{k',l'}(t)\boldsymbol{\alpha}_{n-1}^{k,l}$

Conclusions

- Pros
  - Solution method can answer to the difficulties which the continuous state space creates
    - *Information vector is not discrete*
- Cons
  - Calculations of α-vectors might take time, if many possible control actions, information strategies and solutions for long term are needed

References

Corotis, R. B., Hugh Ellis, J., & Jiang, M. (2005). Modeling of risk-based inspection, maintenance and life-cycle cost with partially observable Markov decision processes. Structure and Infrastructure Engineering, 1(1), 75-84.

White, D. J. (1993). Markov Decision Processes. John Wiley & Sons.

Homework

Use the slides of the presentation (notation) and the article of Corotis et. al (2005) (numeric values) to answer the following questions.

1. Explain what $\boldsymbol{q}_2^{1,3}(t+1)$ is. Calculate the $\boldsymbol{q}_2^{1,3}(t+1)$ when $\boldsymbol{q}(t) = (0.85, 0.10, 0.05, 0, 0)$
2. Explain what $\boldsymbol{\alpha}_1^{3,2}$ is. Calculate the $\boldsymbol{\alpha}_1^{3,2}$ vector when $\beta = 0.99$.

Solutions can be calculated either by hand or with Excel, Matlab or R.

Return solutions until Friday 13th of November to jussi.leppinen@aalto.fi.

Need help: Send email or message in Telegram (@jussi_lep)