Dynamic decision making with stopping

Daniel Karlík

The Czech Academy of Sciences, Institute of Information Theory and Automation

daniel@karlik.cz

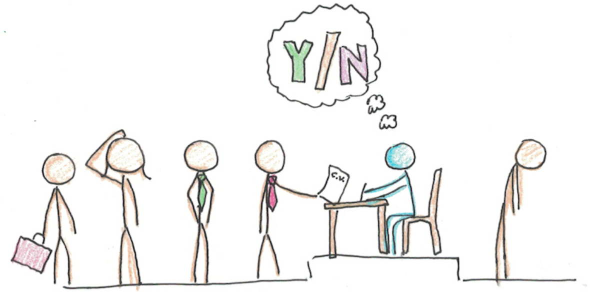
June 23, 2022

Introduction

- Dynamic decision making (DM)
- Stopping rule
- Fully Probabilistic Design

Stopping rule & secretary problem

- Often called stopping time
- As a stopping rule we understand a method for making a decision whether to continue or stop a process
- Famous stopping rule example: **Secretary problem**
- Also known as **marriage problem** or the **best choice problem**



Source ScottHYoung.com Services Ltd.

Fully probabilistic design I.

- Different approach to the regular Markov decision processes
- Markov decision processes use a loss function to evaluate optimal policy
- Whereas FPD quantifies agent's aims and desires probabilistically, from which derives the optimal policy

Fully probabilistic design II.

- We assume a finite set of possible actions $\mathcal{A}$ and possible states $\mathcal{S}$; finite time horizon $|\mathcal{T}|$
- States are denoted as s_t and actions as a_t
- We operate on a closed-loop formed by an agent and its environment
- Probability behaviour of closed-loop:

$$c^{\pi}(b) = \prod_{t \in \mathcal{T}} m(s_t | a_t, s_{t-1}) r(a_t | s_{t-1})$$

- Ideal probability behaviour of closed-loop:

$$c^{i}(b) = \prod_{t \in \mathcal{T}} m^{i}(s_t | a_t, s_{t-1}) r^{i}(a_t | s_{t-1})$$

Fully probabilistic design III.

- Distance between these probabilities is evaluated using Kullback-Leibler divergence

$$D(c^{\pi}||c^{i}) = \int_{b\in\mathcal{B}} c^{\pi}(b)\ln\left(\frac{c^{\pi}(b)}{c^{i}(b)}\right) db$$

- As optimal policy is selected

$$\pi^{o} \in \arg\min_{\pi\in\Pi} D(c^{\pi}||c^{i})$$

- Optimal decision rules can be numerically evaluated using formula:

$$r^{o}(a_t|s_{t-1}) = r^{i}(a_t|s_{t-1})\frac{\exp[-\mathrm{d}(a_t, s_{t-1})]}{\mathrm{h}(s_{t-1})}$$

FPD with stopping I.

We extend the action and state spaces in the following way.

Extension of the classical actions

The actions with stopping are defined as $\alpha_t := (a_t, \tilde{a}_t)$, where a_t represents regular action and $\tilde{a}_t$ is defined as follows

$$\tilde{a}_t := \begin{cases} 1 & \text{continue in generating the regular action } a_t, \\ 0 & \text{take the final regular action } a_t \text{ and stop.} \end{cases}$$

Extension of the classical states

With a similar way to actions we extend the states as $\beta_t := (s_t, \tilde{s}_t)$, where s_t is classical state and $\tilde{s}_t$ reflects if the DM process is stopped or not and is defined as

$$\tilde{s}_t := \begin{cases} 1 & \text{DM process continues,} \\ 0 & \text{DM process is stopped.} \end{cases}$$

FPD with stopping II.

- The key is a proper design of ideal pd c^i
- Process is not stopped $\Rightarrow$ KLD increase with continuing of DM
- Process is already stopped $\Rightarrow$ KLD does not increase
- When we stop we pay for the final choice/selection

Advantages of the use of FPD with stopping

- Most of the hard work is transferred to the design of ideal model, which reflects agent's aims
- This approach seems to be more robust, in can handle changing tasks conditions
- Can extend the static (one-shot) stopping to dynamic DM carring

Open problems

- Constructing of some numerical examples and usages of the proposed method
- Comparison of the results with those obtained by MDP
- Testing this approach on some real world tasks
- Exploitation of possibility to stop a part of dynamic DM

Thank You for your attention

References

T. S. Ferguson (1989)
Who solved the secretary problem?
Statistical Science 4(3), pp. 282 – 296.

Kárný, M., T. V. Guy (2006)
Fully probabilistic control design
Systems & Control Letters 55(6), pp. 259–265.

Kárný, M., T. V. Guy (2014)
On the Origins of Imperfection and Apparent Non-Rationality
Studies in Computational Intelligence 538, pp. 57–92.