Stohastički (Markovljevi) procesi odlučivanja

Adaptivni sistemi – L2

Milan R. Rapaić

Katedra za automatsko upravljanje
Departman za računarstvo i automatiku
Fakultet tehničkih nauka
Univerzitet u Novom Sadu

10. januar 2020

Konačni stohastički procesi odlučivanja

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Co-funded by the
Erasmus+ Programme
of the European Union

Itasdi

Process odlučivanja: Interakcija agenta i okruženja

- AGENT interaguje sa OKRUŽENJEM sekvencijalno, u DISKRETNIM VREMENSKIM TRENUCIMA.
- U svakom diskretnom trenutku vremena t, agent "prima" informaciju o STANJU OKRUŽENJA $s_t \in \mathcal{S}$, i na osnovu te informacije bira način delovanja, odnosno AKCIJU $a_t \in \mathcal{A}$.
- U *sledećem* vremenskom trenutku, okruženje "nagrađuje" agenta NAGRADOM $r_{t+1} \in \mathcal{R} \subseteq \mathbb{R}$.
- Cilj agenta je da maksimizuje DOBIT, tj. određenu kumulativnu meru nagrada koje prima tokom interakcije sa okruženjem.

Process odlučivanja: Interakcija agenta i okruženja

- Pravilo (algoritam, logika) na osnovu koga agent bira narednu akciju poznajući stanje okruženja nazivamo STRATEGIJOM.
- UČENJE je (u ovom kontekstu) postupak promene strategije na osnovu iskustva.
- VREDNOST STANJA s PRI STRATEGIJI u je očekivana dobit ukoliko, polazeći iz stanja s, primenjujemo strategiju u.
- VREDNOST AKCIJE a U STANJU s PRI STRATEGIJI u je očekivana dobit ukoliko, polazeći iz stanja s, u prvim koraku delujemo akcijom a, a nakon toga primenjujemo strategiju u.

Process odlučivanja: Interakcija agenta i okruženja

- VREDNOST STANJA s jeste maksimalna dobit koju agent može ostvariti kada je okruženje u stanju s. Drugim rečima, to je dobit koju agent ostvaruje ukoliko bira najisplativiju strategiju.
- VREDNOST AKCIJE a U STANJU s jeste maksimalna dobit koju agent može ostvariti kada deluje akcijom a ukoliko je okruženje u stanju s.

Konačni stohastički procesi odlučivanja

Konačni stohastički procesi odlučivanja

- Tekuće stanje, tekuća akcija i nagrada koju agent ostvaruje su diskretne slučajne promenljive: S_t, A_t, R_t.
- Jednoznačno je definisana *verovatnoća* da će okruženje koje se nalazi u stanju s kada agent deluje akcijom a preći u stanje s^+ i uzvratiti nagradom r.

$$p(s^+, r|s, a) := \mathbb{P}\left\{S_{t+1} = s^+, R_{t+1} = r|S_t = s, A_t = a\right\}$$

- Dobit agenta u nekom trenutku definisana je kao (otežana) suma svih potonjih nagrada.

$$G_t := \sum_{\tau=0}^{\infty} \gamma^\tau R_{t+1+\tau}$$

Deterministički procesi odličivanja

- Strategija agenta jeste verovatnoća da će agent delovati akcijom a ako je stanje okruženja s.

$$\pi(a|s) := \mathbb{P}\{A_t = a | S_t = s\}$$

- Vrednost stanja s u strategiji π

$$v_\pi(s) = \mathbb{E}\{G_t | S_t = s, \text{ od trenutka } t \text{ agent bira } \pi\}$$

- Vrednost stanja s i akcije a u strategiji π

$$q_\pi(s, a) = \mathbb{E}\{G_t | S_t = s, A_t = a, \text{ od trenutka } t+1 \text{ agent bira } \pi\}$$

Stohastički procesi odlučivanja

Dodatne definicije i neke osobine

- $(\forall s \in \mathcal{S}, a \in \mathcal{A}) \quad \sum_{s^+ \in \mathcal{S}} \sum_{r \in \mathcal{R}} p(s^+, r|s, a) = 1$
- Verovatnoća prelaza stanja

$$p(s^+|s,a) := \mathbb{P}\left\{S_{t+1} = s^+ | S_t = s, A_t = a\right\}$$

$$p(s^+|s,a) = \sum_{r \in \mathcal{R}} p(s^+, r|s,a)$$

- Očekivana nagrada

$$r(s,a) := \mathbb{E}\{R_{t+1}|S_t = s, A_t = a\}$$

$$r(s,a) = \sum_{r \in \mathcal{R}} r\, \mathbb{P}\{R_{t+1} = r|S_t = s, A_t = a\}$$

$$r(s,a) = \sum_{r \in \mathcal{R}} r \sum_{s^+ \in \mathcal{S}} p(s^+, r|s,a)$$