Projekat: Ball and Beam

Nenad Lukić IN 17/2022 i Marija Todorović IN 13/2022

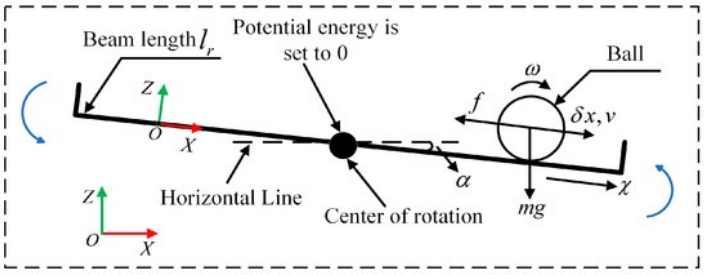
Predmet: Samoobučavajući i adaptivni algoritmi

Fakultet tehničkih nauka - Univerzitet u Novom Sadu

18. februar 2026.

1 Opis problema

Sistem "loptica na gredi" predstavlja klasičan laboratorijski problem. Sistem se sastoji od horizontalne grede čiji se nagib može kontrolisati motorom koji se nalazi na sredini grede i loptice koja se slobodno kotrlja po njoj. Cilj je održati lopticu u željenoj poziciji (na centru) uprkos gravitaciji i brzini loptice, primenom promene ugla nagiba same grede. Glavna ideja jeste da se upotrebom algoritama učenja potkrepljenjem obezbedi zadržavanje loptice na gredi, idealno na sredini. Cilj jeste da obezbedimo da počevši iz bilo kog stanja, o kojima će biti više reči kasnije, agent može da izbalansira samu lopticu. Smatra se da je motor na sredini grede idealan i da greda nema svoj moment inercije. Shema ovakvog sistema je data na slici 1.



Slika 1: Prikaz modela loptice na gredi

Razmatrani sistem se modeluje narednom jednačinom:

$$\ddot{\chi} + \frac{k_f r^2}{mr^2 + J}\dot{\chi} = \frac{mgr^2}{mr^2 + J}\sin\alpha \tag{1}$$

Jednačina se može pronaći na kraju sekcije 2 u narednom radu: https://www.mdpi.com/2073-8994/14/9/1883. Korišćene fizičke veličine zajedno sa njihovim preporučenim vrednostima su:

- m - masa loptice ($0.113\ kg$),
- r - poluprečnik loptice ($0.015\ m$),
- l - dužina grede ($0.4\ m$),
- J - moment inercije za lopticu ($J = \frac{2}{5}mr^2 = 1.017 \times 10^{-5}\ kgm^2$),
- α - ugao rotacije (za $\alpha = 0\ rad$ odgovara ravnotežnom položaju),
- g - gravitaciona konstanta ($g = 9.81\ m/s^2$),
- k_f - koeficijent trenja (0.5 uz pretpostavku da imamo drvenu gredu i metalnu lopticu).

2 Prevođenje matematičkog modela u MPO

Kako bismo mogli da koristimo metode koje su upoznate na ovom predmetu, potrebno je da sam fizički model prevedemo u pogodnu formu. U našem slučaju vreme neće teći kontinualno, već ćemo dati model prevesti u formu koja je ekvivalentna Markovljevom procesu odlučivanja: naredno stanje i nagrada su funkcije tekućeg stanja i primenjene akcije, odnosno:

$$\begin{aligned} s^+ &= f(s,a) \\ r &= h(s,a) \end{aligned} \tag{2}$$

Uvešćemo sledeće smene:

$$\begin{aligned} z_1 &= \chi \\ z_2 &= \dot{\chi} \\ z_3 &= \alpha \\ z_4 &= \dot{\alpha} \end{aligned} \tag{3}$$

nakon čega ćemo uraditi diferenciranje po vremenu leve strane svake jednačine, čime dobijamo sledeći matematički model:

$$\begin{aligned} \dot{z}_1 &= z_2 \\ \dot{z}_2 &= -\frac{k_f r^2}{mr^2+J} z_2 + \frac{mgr^2}{mr^2+J} \sin z_3 \\ \dot{z}_3 &= z_4 \\ \dot{z}_4 &= a \end{aligned} \tag{4}$$

pri čemu je a akcija i $\underline{z} = (z_1, z_2, z_3, z_4)$, tj. u $\underline{z}$ su pobrojane sve promenljive stanja. Sledeći korak je diskretizacija, u našem slučaju koristićemo *Euler 1* diskretizaciju, tj. numeričku integraciju levim pravougaonicima, pošto koristimo nelinearan model. Uvrštavanjem izraza za aproksimiranje prvog izvoda po vremenu u 4, dobijamo:

$$\begin{aligned} z_1^+ &= z_1 + T z_2 \\ z_2^+ &= z_2 + T\left(-\frac{k_f r^2}{mr^2+J} z_2 + \frac{mgr^2}{mr^2+J} \sin z_3\right) \\ z_3^+ &= z_3 + T z_4 \\ z_4^+ &= z_4 + T a. \end{aligned} \tag{5}$$

Akcije će takođe biti diskretizovane i one mogu uzimati naredne vrednosti:

$$a = \ddot{\alpha} \in \{-5 rad/s^2, -2.5 rad/s^2, 0 rad/s^2, 2.5 rad/s^2, 5 rad/s^2\}. \tag{6}$$

Fizički sistem je sad preveden u MPO, uz dobijeni vektor stanja $\underline{z} = (z_1, z_2, z_3, z_4)$, kao i akciju a koja predstavlja ugaono ubrzanje $\ddot{\alpha}$ koje se primenjuje na sredinu grede. Vreme odabiranja je T i njegova vrednost će biti odabrana da bude 0.05 s.

3 Formiranje nagrade

Za formiranje nagrade će biti korišćen DRM (Detail-Reward Mechanism) pristup, što znači da on predstavlja proizvod različitih evaluacionih funkcija s tim da se za svaku od njih smatra da su neprekidno diferencijabilne. U našem slučaju, nagradu definišemo na sledeći način:

$$r = h(s,a) = e_0(s,a) \cdot e_1(s,a) \cdot e_2(s,a)$$

Definicije pojedinačnih evaluacionih funkcija:

1. Kazna za ispadanje (e_0):

$$\Omega_x = \begin{bmatrix} -\frac{l}{2} \le \chi \le \frac{l}{2} \\ -\dot{\chi}_{max} \le \dot{\chi} \le \dot{\chi}_{max} \\ -\alpha_{max} \le \alpha \le \alpha_{max} \\ -\dot{\alpha}_{max} \le \dot{\alpha} \le \dot{\alpha}_{max} \end{bmatrix} \tag{7}$$

$$e_0(s,a) = \begin{cases} 1, & s \in \Omega_x \\ 0, & s \notin \Omega_x \end{cases} \tag{8}$$

Uz sledeće vrednosti: $\dot{\chi}_{max} = 0.5m/s$,$\alpha_{max} = 0.35rad$ i $\dot{\alpha}_{max} = 1rad/s$.

2. Poziciona komponenta (e_1):

$$e_1(s,a) = max\{0, 1 - \frac{2|\chi|}{l}\} \tag{9}$$

Tako da važi sledeće: $e_1 \in [0,1]$.

3. Stabilizacija ubrzanja (e_2):

$$e_2(s,a) = 1 - p \cdot |\ddot{\chi}| \tag{10}$$

Tako da važi sledeće: $e_2 \in [0,1]$ i $p = 1$ je koeficijent kažnjavanja.

4 Postupak rada na projektu

1. Izvršiti diskretizaciju matematičkog modela u prostoru stanja.
2. Definisanje prostora akcija i stanja, s tim da je prostor akcija diskretan.
3. Implementacija načina dodele nagrade.
4. Treniranje agenta primenom Q-Learning i SARSA algoritma uz eksperimentisanje sa različitim vrednostima hiperparametra datih algoritama.
5. Prikaz dobijenih rezultata.
6. Kreiranje pomoćnog alata koji će se koristiti za vizualizaciju.

Literatura

[1] Yao, S.; Liu, X.; Zhang, Y.; Cui, Z. Research on Solving Nonlinear Problem of Ball and Beam System by Introducing Detail-Reward Function. Symmetry 2022, 14, 1883. https://doi.org/10.3390/sym14091883

[2] 4. domaći zadatak: Cart-pole iz predmeta Samoobučavajući i adaptivni algoritmi