به نام خدا

دانشکده مهندسی مکانیک

نام درس: هوش مصنوعی

تمرین ۵(یادگیری تقویتی)

استاد درس: دکتر شریعت‌پناهی

دانشجو:

مهدی نوذری

۸۱۰۶۰۱۱۳۹

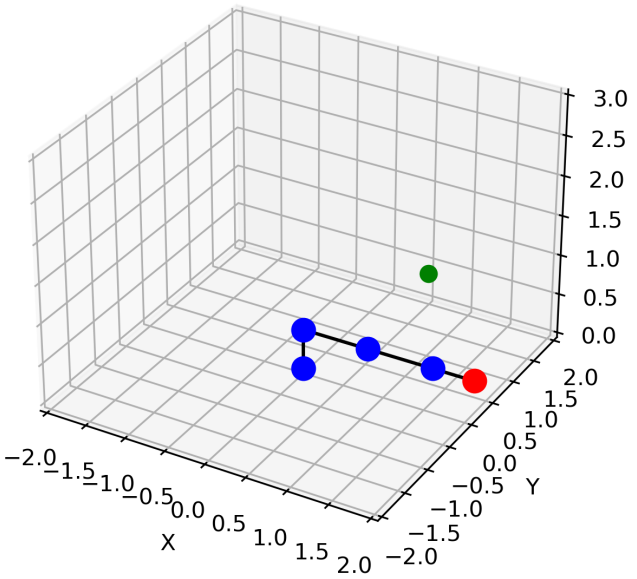
تابستان ۱۴۰۳

تمامی فایل‌ها در Github موجود هستند: https://github.com/Morphit/UT_AI_1403

در این گزارش پیاده‌سازی الگوریتم‌های یادگیری تقویتی بر روی یک بازوی رباتیک انجام می‌شود. در طی این روند یک محیط تعریف شده که ساختار ربات و همجنین محیط اطراف آن را توصیف می‌کند. سپس با تعریف یک عامل به وسیله الگوریتم‌های مختلف، عامل‌ها سعی در پیدا کردن مسیر مناسب برای رسیدن به هدف از یک نقطه اولیه خواهند داشت.

۱ تعریف محیط

محیطی که باید تعریف شود شامل ربات، هدف نهایی و موانع، و همچنین نحوه پاداش‌دهی به عامل می‌باشد. در این تمرین از یک ربات با ساختار مشابه ربات UR3 استفاده می‌شود که دارای ۶ درجه آزادی می‌باشد و تمامی مفاصل آن از نوع Revolute هستند. از جایی که جهت نزدیک شده به هدف توسط بازو اهمیتی ندارد و در این تمرین تنها نیاز است تا انتهای ربات به هدف برسد، می‌توان دو درجه انتهایی ربات را حذف نمود تا در نهایت ربات دارای ۴ درجه آزادی باشد. ساختار انتهایی ربات در شکل ۱ مشخص شده است. برای تعریف این



شکل ۱: ساختار ربات و محیط

ربات به محیط، از جایی که تنها تغییر زاویه و رسیدن به هدف اهمیت دارد، نوشتن سینماتیک مستقیم ربات کافی است. بنابراین سینماتیک

مستقیم این ربات برای هر یک از مفصل‌ها تا انتهای ربات به صورت زیر قابل تعریف می‌باشد.

$$
\begin{aligned}
x &= \cos(\theta_0)\left(L_1\cos(\theta_1) + L_2\cos(\theta_1+\theta_2) + L_3\cos(\theta_1+\theta_2+\theta_3)\right)\\
y &= \sin(\theta_0)\left(L_1\cos(\theta_1) + L_2\cos(\theta_1+\theta_2) + L_3\cos(\theta_1+\theta_2+\theta_3)\right)\\
z &= L_0 + L_1\sin(\theta_1) + L_2\sin(\theta_1+\theta_2) + L_3\sin(\theta_1+\theta_2+\theta_3)
\end{aligned}
\tag{۱}
$$

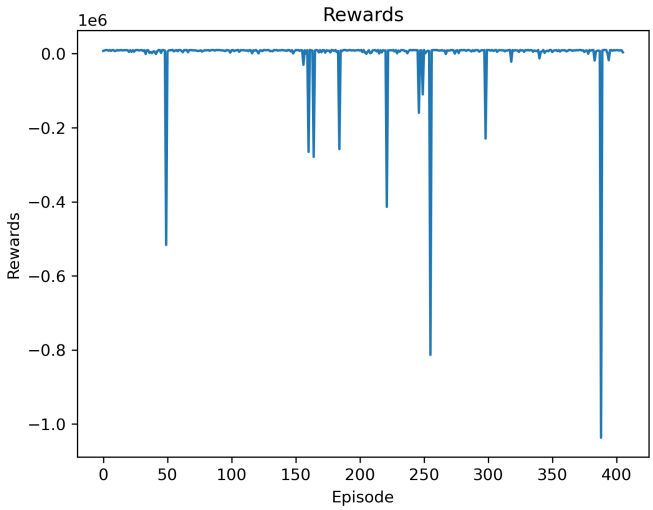
ساختار پاداش‌دهی نیز در جدول ۱ آمده است.

جدول ۱: نحوه پاداش‌دهی به عامل

Action result	**Reward**
Distance to target decreases	+1
Getting close to obstacle	-1000
Reaching target	+10000
Each step	-1

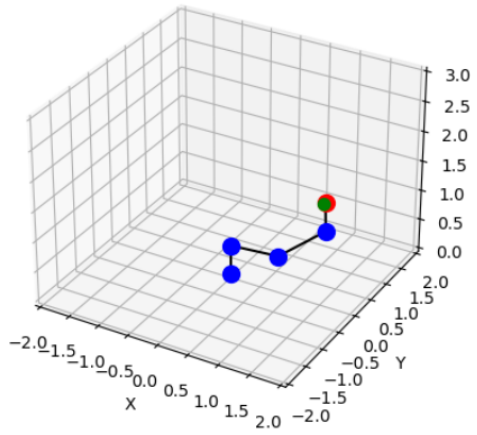
۲ یادگیری Q

حال با سیاست Epsilon-Greedy و با $\epsilon = 0.3$ عامل را تربیت می‌کنیم. تربیت برا ۱۰۰۰ اپیزود انجام می‌شود و در نهایت ۳۸ درصد اپیزودها موفق بوده و نمودار پاداش‌های دریافتی در اپیزودهای موفق در شکل ۲ آمده است.



شکل ۲: مجموع پاداش‌های اپیزودهای موفق

```
Episode 14/1000, Total Reward: 49720, Epsilon: 0.3, steps: 286 , Done: True
Link 1 Joint Angle: 0.7999999999999999
Link 2 Joint Angle: -0.4
Link 3 Joint Angle: 0.7999999999999999
Link 4 Joint Angle: 1.1999999999999997
End Effector: (1.0165632472351274, 1.0466927150336087, 0.9997868015207525)
```



شکل ۳: یک نمونه از اپیزودهای موفق