دانشگاه تهران

دانشکدگان علوم و فناوری‌های میان‌رشته‌ای

درس:

هوش‌مصنوعی

تمرین شماره ۲

مدرس:

دکتر حانیه نادری

دستیاران:

امید استواری

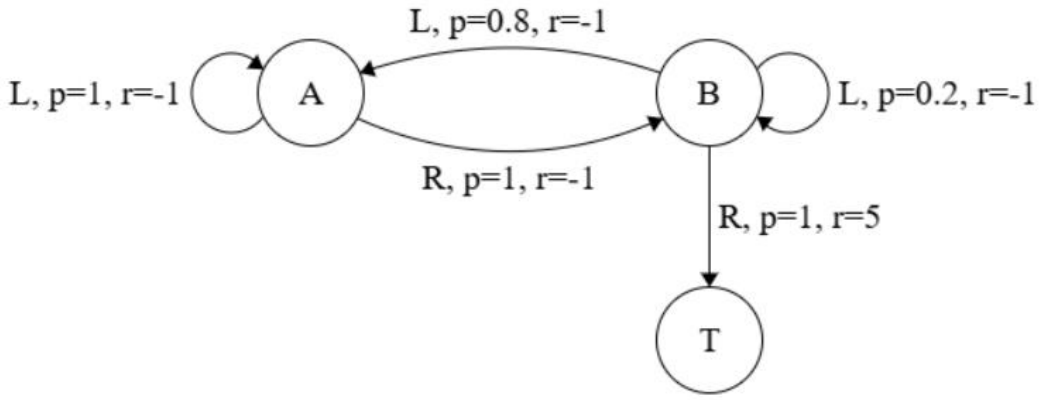
علیرضا میررکنی

بهار ۱۴۰۴

فهرست گزارش سؤالات

سؤال ۱ – دوراهی یادگیری تقویتی

عاملی را در یک محیط با سه حالت A، B و T (که حالت نهایی است) در نظر بگیرید. در حالات A و B عامل می‌تواند بین دو کنش چپ (L) و راست (R) یکی را انتخاب کند. MDP مربوط به این محیط به شکل زیر است.



فرض کنید ضریب تخفیف (discounting factor) برابر $\gamma = 0.9$ است. رسیدن به حالت T منجر به اتمام اپیزود می‌شود.

الف) عناصر یک MDP را نام ببرید، آن‌ها را توضیح دهید و در این محیط این عناصر را تعیین کنید.

ب) فرض کنید π سیاست تصادفی یکنواخت باشد که هر کدام از کنش‌های L و R را به احتمال $\frac{1}{2}$ انتخاب می‌کند. معادله بلمن را برای $V_\pi(A)$ و $V_\pi(B)$ بنویسید و مقادیر عددی آن‌ها را بیابید.

ج) روش value iteration را با شروع از $V_0(A) = V_0(B) = 0$ برای دو مرحله اعمال کنید و $V_2(A)$ و $V_2(B)$ را به دست آورید.

د) از نتیجه قسمت قبل استفاده کنید و سیاست بهینه π^* را تعیین کنید.

ه) عامل انتقال $(A, r = -1, B)$ را مشاهده می‌کند. با فرض اینکه تخمین فعلی به صورت $Q(A,R) = 1.2$ و $Q(B,L) = 0.8$ است، با نرخ یادگیری $\alpha = 0.1$ یک بروزرسانی Q-learning روی $Q(A,R)$ انجام دهید (فرض کنید تمامی Q-value های داده نشده صفر هستند).

سؤال ۲- عامل Q-Learning در CartPole

در این تمرین عملی (فایل CartPole.ipynb) ، هدف شما پیاده‌سازی یک عامل یادگیری تقویتی (Reinforcement Learning Agent) با استفاده از الگوریتم Q-Learning در محیط شبیه‌سازی‌شده‌ی CartPole-v1 از مجموعه محیط‌های Gymnasium است. در این محیط، عامل باید بیاموزد که چگونه با حرکت دادن یک چرخ‌دستی به چپ یا راست، میله‌ای را که روی آن قرار دارد در حالت تعادل نگه دارد. کدی که در اختیار شما قرار گرفته، ساختار کلی پروژه را فراهم کرده اما بخش‌هایی از آن به صورت TODO # مشخص شده‌اند تا شما آن‌ها را تکمیل کنید.

این بخش‌ها شامل طراحی تابعی برای ایجاد بین‌ها به منظور گسسته‌سازی فضای مشاهدات پیوسته، پیاده‌سازی تابعی برای تبدیل وضعیت‌های پیوسته به حالت‌های گسسته با استفاده از np.digitize، نوشتن منطق انتخاب عمل بر اساس سیاست ε-greedy، پیاده‌سازی به‌روزرسانی جدول Q بر اساس معادله بلمن، تکمیل حلقه آموزش عامل شامل کاهش تدریجی مقدار ε و محاسبه پاداش کلی در هر اپیزود، ترسیم نمودار میانگین متحرک پاداش برای بررسی روند یادگیری، و همچنین نوشتن کدی برای ارزیابی عملکرد نهایی عامل در محیط با ضبط فریم‌ها و ساخت فایل GIF از اجرای نهایی عامل می‌باشد.

پس از اجرای کامل این تمرین، عامل آموزش‌دیده‌ی شما باید قادر باشد میله را برای مدت مناسبی در حالت تعادل نگه دارد و عملکرد آن به صورت بصری در فایل cartpole_result.gif و نمودار پاداش قابل مشاهده باشد. این تمرین به شما کمک می‌کند تا درک عمیق‌تری از مفاهیم اصلی یادگیری تقویتی، از جمله گسسته‌سازی فضاهای پیوسته، سیاست‌های انتخاب عمل و پیاده‌سازی الگوریتم‌های یادگیری مبتنی بر Q داشته باشید.

نکات

- مهلت تحویل تمرین ۲۹ اردیبهشت ۱۴۰۴ است.
- انجام این تمرین به‌صورت یک نفره می‌باشد.
- حداکثر مهلت مجاز برای تأخیر تمرین‌ها ده روز خواهد بود (دقیقاً ۱۰ روز پس از مهلت آپلود سامانه بسته خواهد شد).
- گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است. لطفاً تمامی نکات و فرض‌هایی که برای پیاده‌سازی‌ها و محاسبات خود در نظر می‌گیرید را در گزارش ذکر کنید.
- کدهای خود را به‌صورت عکس در داخل گزارش کپی نکنید و با فرمت مناسب آن را در گزارش قرار دهید.
- داخل کدها کامنت‌های لازم را قرار دهید و تمامی موارد مورد نیاز برای اجرای صحیح کد را ارسال کنید.
- الزامی به ارائه توضیح جزئیات کد در گزارش نیست اما باید نتایج به‌دست آمده را گزارش و تحلیل کنید.
- گزارش را در قالب تهیه شده که روی صفحه درس در سامانه eLearn بارگذاری شده بنویسید.
- در گزارش خود برای تصاویر زیرنویس و برای جداول هم بالانویس اضافه کنید.
- اگر بخشی از کد را از کدهای آماده اینترنتی استفاده می‌کنید که جزء قسمت‌های اصلی تمرین نمی‌باشد، حتماً باید لینک آن در گزارش و کد ارجاع داده شود در غیر این صورت تقلب محسوب شده و کل نمره تمرین را از دست می‌دهید ولی محدودیتی در استفاده از منابع اینترنتی ندارید.
- لطفاً فایل کدها و سایر ضمائم مورد نیاز را با فرمت زیر در صفحه درس در سامانه eLearn بارگذاری نمایید.
HW1_[Lastname] _ [StudentNumber].zip

- در صورت وجود هرگونه ابهام یا مشکل می‌توانید از طریق رایانامه‌های زیر یا تلگرام با دستیاران آموزشی طراح تمرین در تماس باشید.
- طراح سؤال‌ها:

علیرضا میرکنی: alirezamirrokni28@gmail.com

- در صورت مشکل در آپلود در سامانه درس یا مشکلاتی از این قبیل می‌توانید با دستیار ارشد درس در ارتباط باشید:

امید استواری: omid.ostovari@ut.ac.ir