



آموزش موثر دانش‌آموزان و دانشجویان در مدارس و دانشگاه‌ها، یکی از مهم‌ترین چالش‌های موجود در دنیا است. در بحث آموزش موثر، ارائه یک محتوای مناسب به یادگیرندگان که براساس آن آموزش ببینند، از اهمیت بالایی برخوردار است. یک موسسه که تعدادی از مدارس کشور را مدیریت می‌کند، تصمیم گرفته برای حل این مشکل از روش‌های مبتنی بر **Multi-armed bandit** استفاده کند. این موسسه می‌خواهد در ۱۰ مدرسه و به صورت مستقل سیستم خود را برای یافتن بهترین برنامه آموزشی اجرا کند. توزیع دانش‌آموزان از لحاظ ویژگی‌های متفاوت در این ۱۰ مدرسه مشابه می‌باشد. این موسسه ۳ برنامه آموزشی مختلف را تهیه کرده است. هر برنامه برای اجرا در یک هفته آموزشی بوده و مطلوب بودن یک برنامه، در انتهای هفته و براساس عملکرد دانش‌آموزان ارزیابی می‌شود. با فرض اینکه هر مدرسه دارای ۱۰۰ دانش‌آموز پایه دهم می‌باشد که این سیستم برای آن‌ها اجرا می‌شود و هر هفته به هر دانش‌آموز یک برنامه خاص پیشنهاد داده می‌شود، به سوالات زیر پاسخ دهید:

(۱) مسئله را به صورت یک مدل **Multi-armed bandit** مدل کنید. بازوها و تابع پاداش را مشخص کنید

(۲) با توجه به کدهایی که در اختیار شما قرار داده شده است، کلاس مربوط به عامل یادگیرنده مبتنی بر الگوریتم **gradient bandit** را پیاده‌سازی کنید. در ادامه باتوجه به توابع عملکرد دانش‌آموزان و مدل محیط، با کمک عامل یادگیرنده و در مدت زمان یکسال، مشخص کنید بهترین برنامه آموزشی کدام برنامه است. عامل یادگیرنده را نرخ یادگیری ۰.۰۰۱، ۰.۰۱ و ۰.۱ اجرا کرده و با رسم نمودار **reward** و **regret** عملکرد آن‌ها را مقایسه کنید

(۳) مدل‌های زبانی بزرگ^۱ یکی از ابزار مبتنی بر هوش مصنوعی هستند که می‌توانند در زمینه‌های مختلفی کمک‌کننده باشند. یکی از حوزه‌هایی که می‌توان از این مدل‌ها استفاده کرد، برای حل مسائل یادگیری تقویتی است (مطالعه مقالات ارائه شده توصیه می‌شود). در این قسمت لازم است بدون استفاده از الگوریتم‌های یادگیری تقویتی و تنها با استفاده از مدل‌های زبانی بزرگ مسئله بیان شده در بالا را حل کنید. در این قسمت باید مسئله را برای مدل زبانی مطرح کرده و پاسخی را از آن برای حل مسئله دریافت کنید. این پاسخ‌ها را بر اساس دانش خود ارزیابی کرده و متکی بر آن مسئله را حل کنید. هدف از انجام این قسمت این می‌باشد که بررسی کنید آیا می‌توان از دانش یاد گرفته شده توسط مدل‌های زبانی بزرگ برای بهبود روند حل مسئله استفاده کرد؟ با تحلیل پاسخ‌های مدل این مورد را بررسی کنید.

(۴) فرض کنید مسئولین اجرای این سیستم در یکی از مدارس، می‌توانند عملکرد اجرایی در ۹ مدرسه دیگر را مشاهده کرده و بدانند که در هر هفته به دانش‌آموزان مدارس دیگر چه برنامه‌های آموزشی پیشنهاد داده می‌شود، اما از مطلوب بودن عملکرد این مدارس و برنامه پیشنهادی آنان اطلاعی نداشته باشند. با استفاده از مدل‌های بزرگ زبانی مشخص کنید دانستن این اطلاعات اضافی به مسئولین اجرایی در این مدرسه خاص چه کمکی می‌کند؟ آیا آنان می‌توانند زودتر برنامه پیشنهادی مطلوب را برای دانش‌آموزان خود شناسایی کنند؟ چگونه؟

(۵) مشکلی که در حل مسئله در قسمت اول و دوم وجود دارد، این است که برای همه دانش‌آموزان با ویژگی‌های متفاوتی که دارند یک مدل را حل کرده و یک بازو را که به طور متوسط بهتر بود انتخاب می‌کردیم. برای حل چنین مسائلی می‌توان از نوعی از **multi-armed bandit** ها تحت عنوان **contextual bandit** استفاده کرد. در مورد این مدل‌ها تحقیق کرده و ویژگی‌های آن را بیان کنید. در این مسئله می‌توان دانش‌آموزان را باتوجه به اینکه بیشتر از طریق محتوای گفتاری و نوشتاری، محتوای گرافیکی و یا انجام تجربه عملی یادگیری بهتری دارند، به ۳ دسته تقسیم کرد. همچنین هر مدرسه از هر دانش‌آموز ۳ ویژگی را می‌شناسد که در جدول یک نشان داده شده است. با توجه به این اطلاعات و براساس کدهایی که در اختیار شما قرار گرفته که در آن دانش‌آموزان ۱۰ مدرسه با ویژگی‌هایی که دارند مشخص شده‌اند، راحل را به گونه‌ای پیاده‌سازی کنید که به ازای هر نوع از دانش‌آموزان بهترین برنامه آموزشی شناسایی شود. عملکرد عامل (عوامل) یادگیرنده را

¹ Large language models



با رسم نمودارهای reward و regret بررسی کنید. به نظر شما بجز روشی که در این قسمت بیان شد، از چه روش‌های دیگری برای حل این نوع مسائل می‌توان استفاده کرد؟

جدول 1 - ویژگی‌هایی که هر مدرسه از هر دانش‌آموز خود می‌داند

ویژگی اول	در امتحانات قبلی چند درصد پاسخ‌های دانش‌آموز به صورت متن نوشتاری بوده (بقیه پاسخ‌ها به صورت نمودار و شکل بوده است)
ویژگی دوم	در چند درصد مواقع دانش‌آموز برای شرکت در فعالیت‌های آزمایشگاهی فوق برنامه داوطلب شده است
ویژگی سوم	در چند درصد مواقع انجام بحث کلاسی را به مشاهده ویدیویی مربوط به موضوع در کلاس‌های درس ترجیح داده است

مقالات بخش سوم

- [Spring: GPT-4 Out-Performs RL Algorithm by Studying Papers and Reasoning](#)
- [Reasoning with Language Model is Planning with World Model](#)

اهداف تمرین

تمرین ارائه شده با هدف توسعه مهارت‌های گوناگون ارائه شده است. این تمرین از طریق مواجهه با مسئله‌ای از مسائل دنیای واقعی و مدل‌سازی آن، به دنبال تقویت توانایی حل مسئله و تفکر انتقادی است. علاوه بر این، با تأکید بر تقویت مهارت‌های ارتباطی و کار گروهی، این تمرین فرصت همکاری بین دانشجویان را فراهم می‌کند تا از دیدگاه‌های یکدیگر برای حل مسائل بهره‌مند شوند. به کارگیری خلاقیت در حل مسائل نیز از دیگر اهداف این تمرین است. همچنین استفاده از ابزار روز مانند مدل‌های بزرگ زبانی و بررسی توانایی و امکان کمک آن‌ها در حل مسائل، از جمله مسائل یادگیری تعاملی، از دیگر اهداف این درس می‌باشد.

نکات تمرین

- استفاده از LLM ها در این تمرین مشکلی ندارد. اما در صورت استفاده لطفاً منبع و prompt خود را ذکر نمایید تا ثقل محسوب نشود.
- مهلت ارسال این تمرین تا پایان روز چهارشنبه ۸ آذر خواهد بود.
- انجام این تمرین به صورت گروهی می‌باشد. همچنین بحث و گفت‌وگو در دیسکورد مانعی ندارد.
- لطفاً گزارش و کد تمرین را در فایل‌هایی که از طریق google Doc و google colab با شما به اشتراک گذاشته شده است، وارد نمایید.
- در صورت وجود سؤال و یا ابهام می‌توانید در channel مربوط به این تمرین با دانشجویان دیگر مطرح نمایید و یا برای ارتباط با دستیاران آموزشی از طریق ایجاد یک thread در همان channel دیسکورد، سؤال خود را مطرح نمایید.