دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده برق و کامپیوتر

پروژه درس یادگیری تعاملی

پاییز 1401

امیرعلی سلطانی - امیرحسین بیرژندی

...

810198367 - 810100376

...

فهرست

چکیده

در این پروژه کوشش می‌کنیم از مفاهیم یادگیری تعاملی برای شخصی‌سازی آموزش در EdTech استفاده کنیم. ابتدا به بررسی اهمیت شخصی‌سازی می‌پردازیم سپس راهکار هایی را برای شخصی سازی با استفاده از یادگیری تعاملی ارائه می‌دهیم.

مقدمه

یکی از تکنولوژی های داغ حال حاضر دنیا EdTech می‌باشد. EdTech به هر گونه تکنولوژی‌ای گفته می‌شود که به منظور درگیر کردن بیشتر و شخصی سازی یادگیرنده‌ها استفاده می‌شود.

از آنجایی که نیاز روز افزون آموزش حس می‌شود و مفهوم هایی همچون lifetime learning و ... مطرح می‌شوند دیگر صرف کلاس های حضوری مدرسه و دانشگاه جوابگوی آموزش نخواهند بود. به همین دلیل نیاز داریم از تکنولوژی برای تسهیل فرآیند آموزش استفاده کنیم.

نکته قابل توجه دیگر که نظر کاربران را برای استفاده از EdTech جلب می کند این است که در کلاس های کلاسیک با حضور تعداد قابل توجهی یادگیرنده، معلم قابلیت تدریس مطابق با سرعت و سلایق هر فرد را ندارد و با توجه به میانگین کلاس تدریس خواهد کرد. در نتیجه اگر بتوانیم با هوش مصنوعی و الگوریتم های یادگیری تعاملی ویژگی های افراد را بشناسیم قادر خواهیم بود محتوای آموزشی را برای هر فرد با توجه به ویژگی های شخصی او فراهم کنیم.

مساله‌ای که قرار است در این بخش حل شود، مدلی است که با توجه به آن می‌توان شخصی‌سازی در آموزش را به‌گونه‌ای وسعت بخشید که با توجه به ویژگی‌های فردی، اجتماعی و محیط درسی بتوان بسته‌ای را به شخص ارائه کرد که با استفاده از این بسته‌ها بتواند پاداش خود در طول زمان را بهینه سازد.

بخش 1 – سیستم توصیه‌گر آموزشی

در این بخش سیستم توصیه گری طراحی می‌کنیم که یاد می‌گیرد برای هر گروه از مشتریان(یادگیرندگان) بسته آموزشی پیشنهاد دهد که آن افراد نمره بیشتری دریافت کند.

فرمول‌بندی:

برای حل این مسئله ما یک Contexual Bandit در نظر گرفته‌ایم که با دریافت پاداش، یک بسته آموزشی توصیه می‌شود.

$$contexts = \{Primary\ School\ Student, High\ School\ Student, University\ Student, Middle\ Age\ Learner\}$$

در این بخش پاداش ما صرفا نمره‌ای است که فرد در ازای استفاده از بسته پیشنهادی گرفته است.

$$reward = Grade$$

اعمال ما نیز در این بخش صرفا بسته آموزشی‌ای است که به فرد ارائه می‌دهیم.

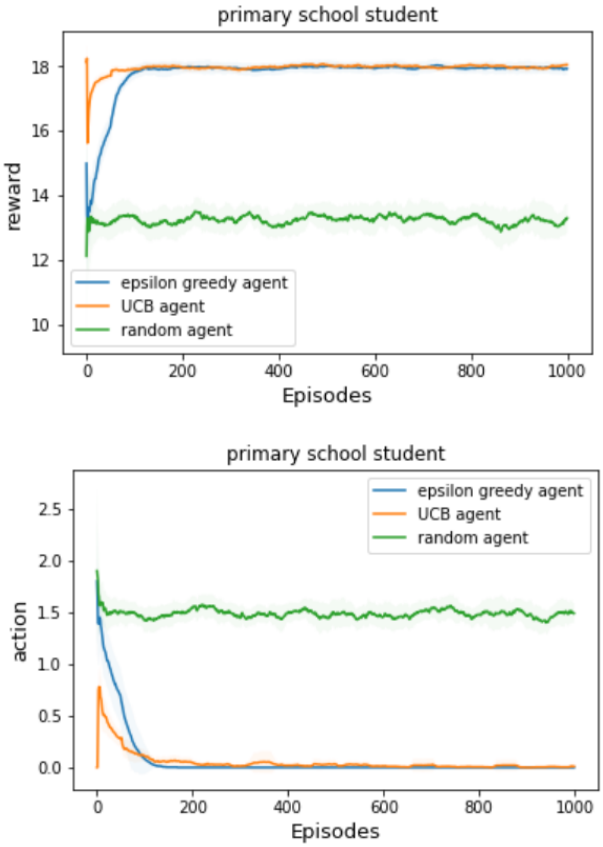
$$actions = \{Group\ Class, Video, Private\ Class, PowerPoint\}$$

نتایج:

برای حل این مسئله از 3 الگوریتم e-greedy، UCB و رندوم استفاده می‌کنیم.

$Primary\ School\ Student$:

$Group\ Class \sim N(11,1)\ , Video \sim N(3,1), Private\ Class \sim N(10,1), PowerPoint \sim N(1,1)$



همانطور که مشاهده می‌کنیم دو الگوریتم e-greedy و UCB به اکشن صفر که همان کلاس گروهی است همگرا شده است و همچنین نمره 18 همگرا شده است.

بخش 2 – سیستم توصیه‌گر آموزشی شخصی‌سازی شده

در این بخش کار را کمی فراتر می‌ببریم و بسته آموزشی را با درنظر گرفتن ویژگی شخصیتی و بودجه فرد پیشنهاد می‌دهیم. به عبارتی هر بسته آموزشی خود یک قیمت دارد و ما بهترین بسته را پیشنهاد خواهیم داد.

فرمول‌بندی:

این مسئله نیز یک Contexual Bandit است که با دریافت پاداش که خود با توجه به ویژگی شخصیتی و بودجه فرد، یک بسته آموزشی توصیه می‌شود.

$$contexts = \{Primary\ School\ Student, High\ School\ Student, University\ Student, Middle\ Age\ Learner\}$$

ویژگی شخصیتی افراد را با یک بردار با دو درایه به صورت زیر مدل می‌کنیم. درایه اول اهمیت یادگیری را برای فرد نشان می‌دهد و درایه دوم در واقع اهمیت پول باقی مانده را در نظر می‌گیرد.

$$\theta = \begin{bmatrix} Learning\ Rate \\ Financial\ Rate \end{bmatrix}$$

حال نحوه محاسبه پاداش به صورت زیر خواهد بود.

$$reward = \theta_0(Grade) + \theta_1(\frac{budget - cost}{budget})$$

اعمال ما نیز کمی با بخش قبل تفاوت دارد و بسته های ترکیبی نیز داریم.

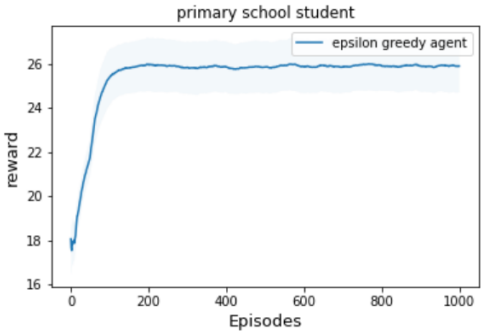
$$action = [Group\ Class, Video, Private\ Class, PowerPoint, Group\ Class\ \&\ Video,$$
$$Group\ Class\ \&\ Private\ Class, Group\ Class\ \&\ PowerPoint, Video\ \&\ Private\ Class,$$
$$Video\ \&\ PowerPoint, Private\ Class\ \&\ PowerPoint]$$

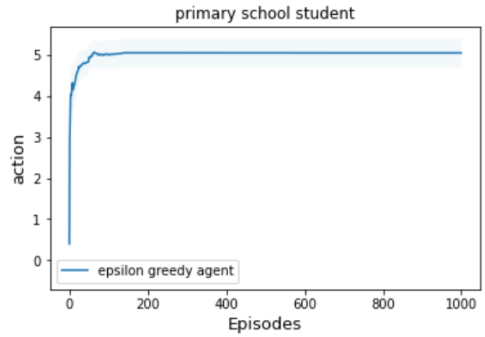
هزینه بسته ها نیز به صورت زیر می‌باشد.

$$costs = \{60, 40, 120, 20, 100, 180, 80, 160, 60, 140\}$$

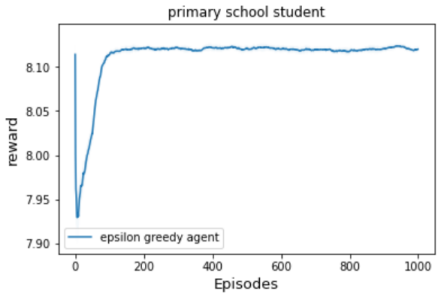
نتایج:

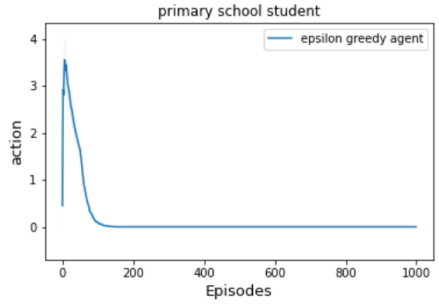
$Primary\ School\ Student, \theta = [1,0], budget = 200$



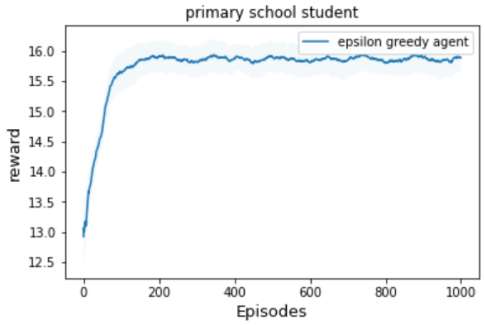


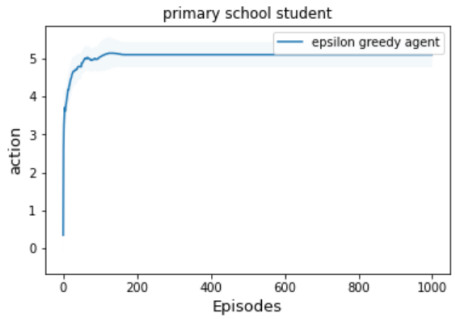
$Primary\ School\ Student, \theta = [0.05,0.95], budget = 150$



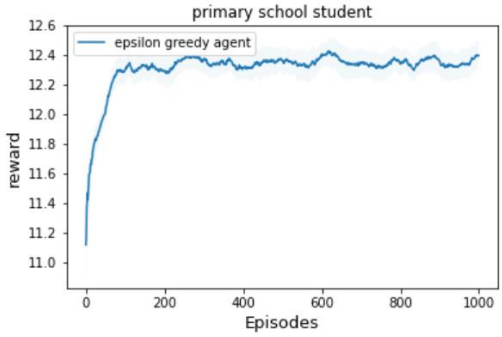


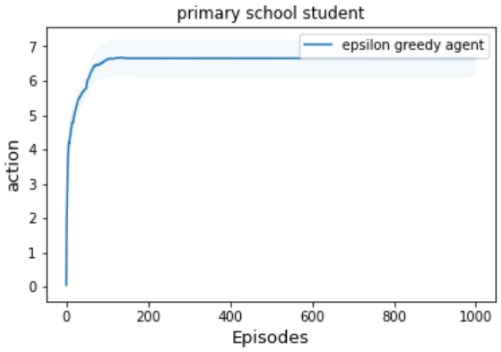
$High\ School\ Student, \theta = [0.6,0.4], budget = 200$





$University\ Student, \theta = [0.5,0.5], budget = 80$





بخش 3 – سیستم توصیه‌گر آموزشی بر مبنای ویژگی

فرمول‌بندی

ابتدا می‌بایست توضیح داد که برای هر عامل، ویژگی‌هایی مبنی بر ویژگی‌های ثابت و متغیر توصیف شده است. ویژگی‌های ثابت هر شخص شامل نوع مدرسه، جنسیت، سطح تحصیلات پدر و مادر، سن، IQ، هوش اجتماعی و تست شخصیتی MBTI در نظر گرفته شده است. ویژگی‌های متغیری که می‌تواند بر روی وی تاثیر بذارد نیز شامل زمان مطالعه وی در مدرسه، زمان آزاد که در طول هفته دارد، زمان مطالعه بسته آموزشی و نوع بسته‌ای است که استفاده می‌کند. در کل برای هر یک از این ویژگی‌های ثابت توزیع نرمالی با توجه به اهمیت هر کدام در دانش و پاداش اجتماعی تعریف شده است. برای زمان‌های مطالعه و آزاد نیز با توجه به هر کدام پاداشی متناسب با زمان اختصاص گرفته توصیف شده است. در کل با توجه به این ۱۱ ویژگی و رابطه مربوط به هر کدام از این پاداش‌ها، پاداش اجتماعی تعریف شده و مقداری جهت بدست آوردن نمره تعریف شده است. با توجه به میزان اختلاف نمره شخص با baseline که مقدار قبولی را ۷۰ از ۱۰۰ گرفته است، پاداش نهایی محاسبه شده است.

با استفاده از این‌که در این روش از مدل armed-bandit استفاده شده است می‌توان گفت که برای هر اکشن، می‌توان توزیعی را توصیف کرد. هر اکشن در این بخش شامل بسته‌ی آموزشی‌ای است که می‌توان به هر شخص ارائه داد. اکشن‌ها هر کدام به شکل زیر توصیف می‌شوند. اکشن اول را استفاده از بسته کلاس‌های گروهی که رابطه مربوط به پاداش این اکشن به شکل زیر توصیف می‌شود.

$$R_{edu} = \frac{t}{10} \times N(10,1) \qquad R_{emo} = \frac{t}{10} \times N(0,1)$$

برای اکشن بعدی که استفاده از کلاس خصوصی هست، رابطه پاداش به شکل زیر توصیف می‌شود.

$$R_{edu} = \frac{t}{9} \times N(10,1) \qquad R_{emo} = \frac{t}{9} \times N(-5,1)$$

برای اکشن سوم که مربوط به ویدیوهای آموزشی هست نیز پاداش‌ها به صورت زیر تعریف می‌شوند.

$$R_{edu} = \frac{t}{8.5} \times N(8,1) \qquad R_{emo} = \frac{t}{10} \times N(0,1)$$

برای اکشن چهارم که نیز مربوط به محتواهای متنی است می‌توان پاداش‌ها را به صورت زیر توصیف کرد.

$$R_{edu} = \frac{t}{10} \times N(10,1) \qquad R_{emo} = \frac{t}{7} \times N(7,1)$$

متد

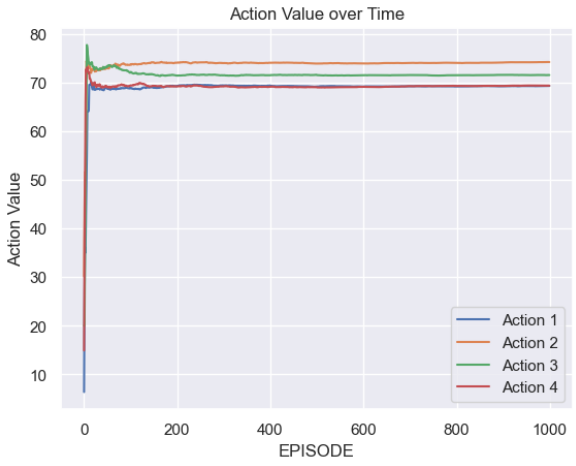
متدی که در این روش استفاده شده است یک مدل armed-bandit است که چهار اکشن را به صورت ورودی می‌گیرد. در این متد با توجه به پاداش‌های توصیف شده فرآیند صورت گرفته است. این مدل armed-bandit نیز با استفاده از روش epsilon-greedy حل شده و حالت‌های مربوط به exploration و exploitation نیز بدست آمده‌اند. برای این‌که بتوان ارزش نهایی را نیز در بازه بین ۰ تا ۱۰۰ بدست آورد نیز ابتدا مقادیر مربوط به هر کدام از ارزش‌ها در مقیاس ۰ تا ۱ با استفاده از رابطه زیر نرمالیزه شده‌اند.

$$x_{norm} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

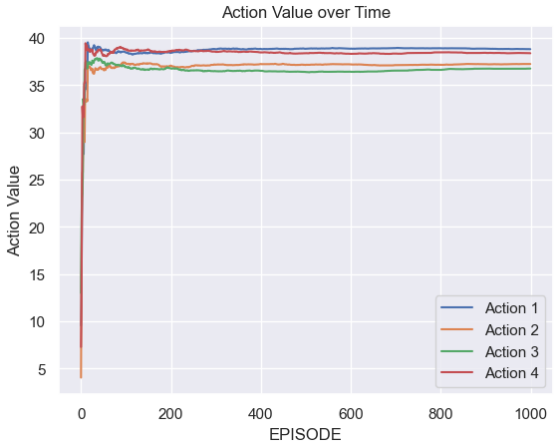
سپس با استفاده از تابعی که به عنوان utility برای هر شخص تعریف شده است، ارزش نهایی با استفاده از ضرایب محاسبه شده‌اند. برای هر کلاس نیز برای این‌که بتوان ارتجاع را بوجود آورد نیز، زمان به صورت مساوی بین حالت‌های استراحت و مدرسه تقسیم شده است.

نتیجه

در حالتی که مقدار utility را تغییر داده شده است، نمودارهای زیر بدست آمده‌اند.

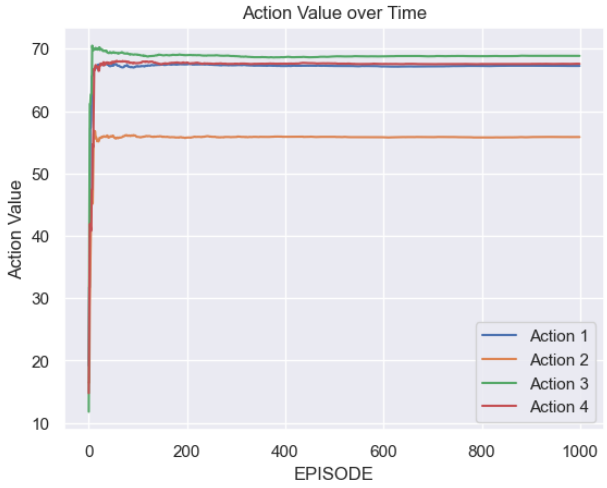


شکل: در حالتی که **p_edu=1** است.



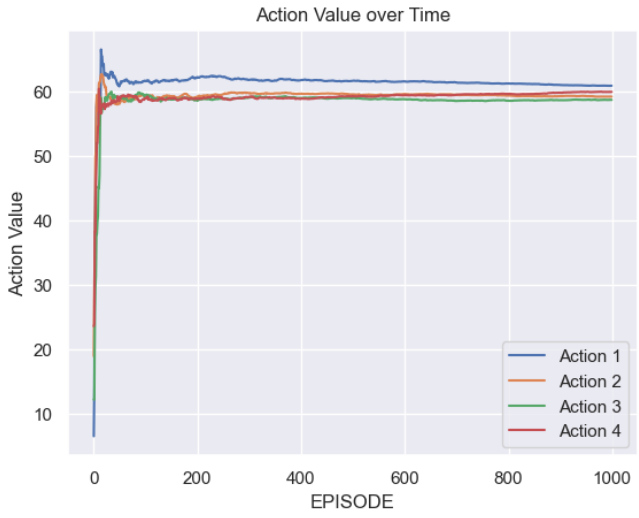
شکل: در حالتی که **p_edu=0.8** است.

شکل: در حالتی که **p_edu=0.7** است.

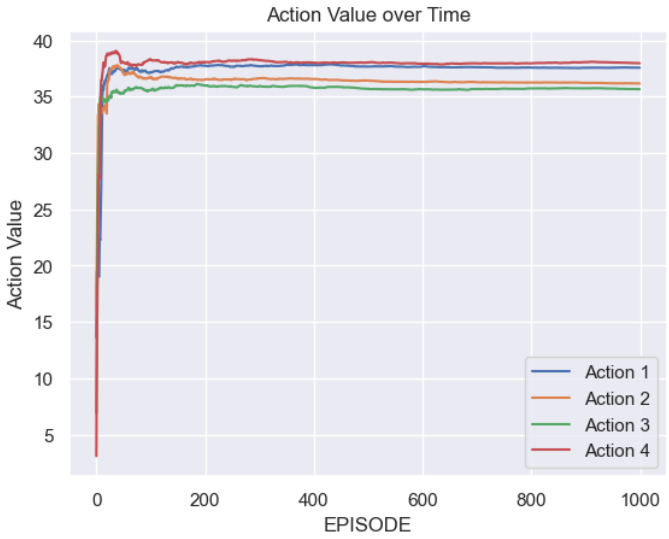


شکل: در حالتی که **p_edu=0.5** است.

همچنین اثر هر کدام از ویژگی‌های ثابت نیز در شکل‌های زیر قابل مشاهده است که در هر کدام از این شکل‌های مقدار utitlity برابر در نظر گرفته شده است.



شکل: شخص اول



شکل: شخص دوم

همان‌طور که مشاهده می‌شود با اینکه utitlity فرق دارد ولی اکشن پیشنهادی تفاوت دارند.

- Reinforcement Learning in Education: A Multi-Armed Bandit Approach

- Student Intervention System using Machine Learning Techniques