

Assignment One

1. تعداد حالت ها وابسته به ابعاد بازی است. میتوان خانه های بلاک را حذف کرد

2. State: مختصات خانه ای که روی آن هستیم

Action: حرکت به سمت بالا راست چپ پایین

Reward: اگر به پرچم برسیم جایزه 50 میگیریم

اگر به تارگت برسیم جایزه $1 - \frac{\text{len(flags)}}{1}$ میگیریم

اگر به پرچم برسیم جایزه $\frac{\text{len(flags)}}{1}$ میگیریم

به ازای هر حرکت $\frac{\text{boardSize}}{1}$ مجازات میشویم

به ازای حرکت به خانه ای که تا به حال در آن بوده ایم $\frac{\text{boardSize}}{2}$ مجازات میشویم

3. گاما نرخ یادگیری را تعیین میکند. اگر گاما صفر باشد جایزه صرفا به انتخاب ما بستگی دارد و مهم نیست به چه state ای میرویم.. و اگر بینهایت باشد دیگر مهم نیست چه action ای انجام میدهیم.

آلفا نرخ یادگیریست.. اگر صفر باشد q-table بدون تغییر باقی میماند و اگر 1 باشد بدون توجه به یادگیری های گذشته q-table را آپدیت میکند.

حالا نتیجه را برای مقادیر مختلف بررسی می کنیم.

برای مقادیر $\alpha = 0.25, 0.5, 0.75$ زمانی که آموزش را شروع می کنیم هرچه α را کمتر می کنیم تعداد iteration های بیشتری طول میکشد تا q-table به راه حل برسد. اشتباهات بیشتری انجام میدهد اما از اشتباهاتش درس میگیرد. برای $\alpha = 0.75$ در اکثر مواقع با وجود آپدیت سریع مسیرها اما بعد از چند iter منحرف میشود و دوباره برمیگردد.

برای مقادیر $\gamma = 0.25, 0.5, 0.75$ بیش

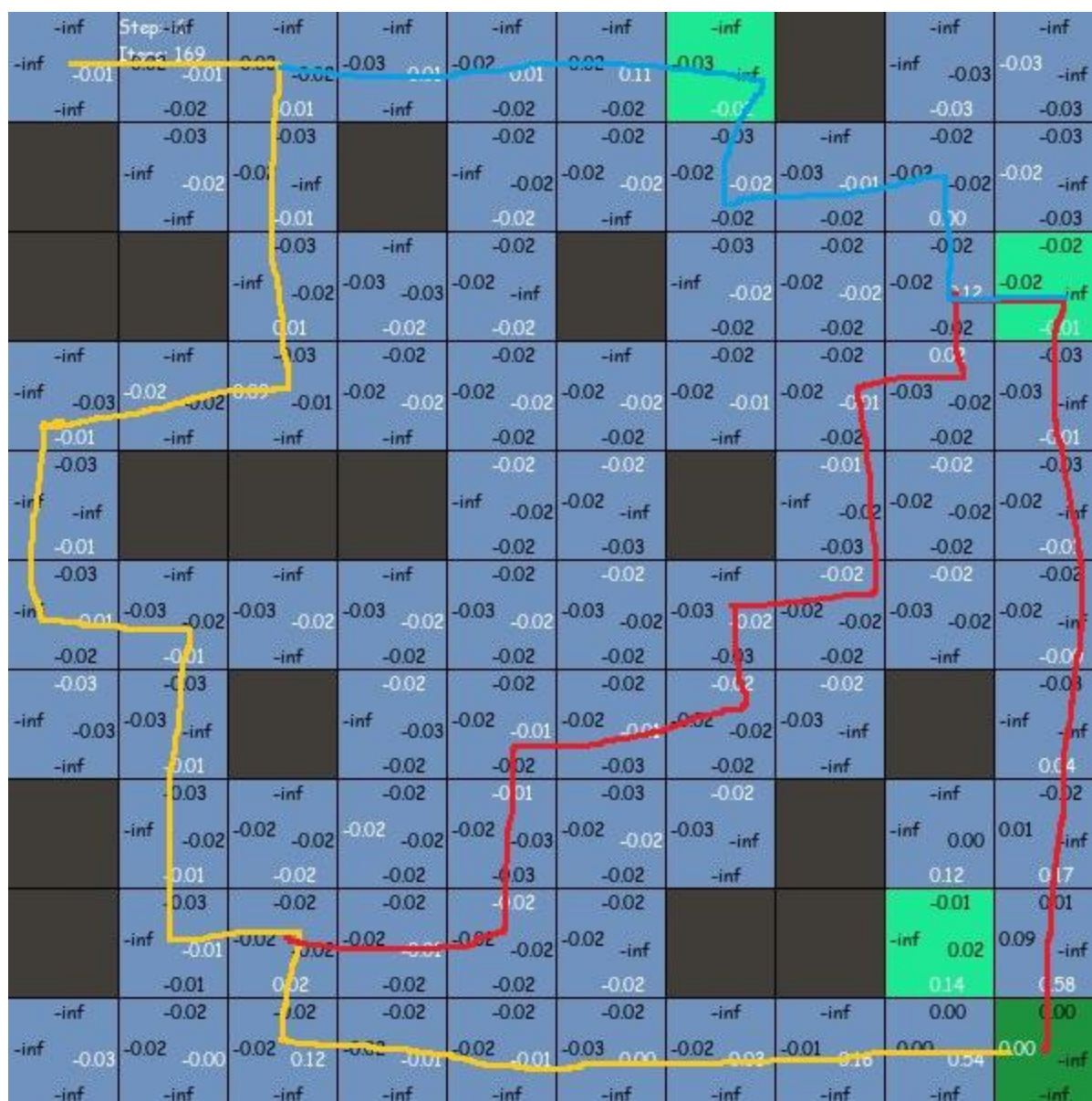
4. میتوانیم با تابع `printQtable` در `main` جدول q هارا چاپ کنیم.

تغییرات reward در زیر آمده است. رنگ سفید بیشترین reward یک state است

5. ابتدا با تابع `generate_random_env()` در `main` یک محیط رندوم ایجاد می کنیم. سپس تابع `train_random()` را فراخوانی می کنیم.

مثال:

با انجام چند مثال از محیط های رندوم متوجه میشویم مدل در رسیدن به `flag` هایی که در مسیر رسیدنش به `target` نیست یا از بقیه `flag` ها دور افتاده ناتوان است. برای مثال `board` زیر را در نظر بگیریم



مدل در مسیرهای مختلف سعی میکند flag هارا بگیرد اما موفق به گرفتن تمام آنها نمی شود.

مشکل کاملاً اساسی در گرفتن flag هایی است که در بن بست قرار دارند چون بعد از گرفتن آنها باید برگردیم.. اما با بازگشت reward منفی میگیریم . دوباره به بن

بست وارد میشویم تا **flag** را بگیریم اما **flag** ای وجود ندارد. انقدر اینکار را انجام میدهیم تا دیگر هرگز وارد بن بست نشویم.