سید علی موسوی رضا فرخ

سوال ۱

با اجرای الگوریتم policy iteration با پارامترهای های گفته شده در custom_map_1:

برای مقدار discount_factor برابر 1 مسئله در یک لوپ میوفتد، زیرا مقدار دلتا کاهش پیدا نمیکند و همواره بزرگتر از تتا است، پس شرط break اتفاق نمیوفتد.

برای بقیه مقادیر discount_factor نتیجه یکسان است، زیرا محیط به گونه ای است که، مقدار discount_factor تاثیر خاصی نمیگذارد. (به دلیل سادگی محیط)

سوال ۲

با اجرای الگوریتم policy iteration با پارامترهای های گفته شده در custom_map_2:

برای مقدار discount_factor برابر 1 مسئله در یک لوپ میوفتد، زیرا مقدار دلتا کاهش پیدا نمیکند و همواره بزرگتر از تتا است، پس شرط break اتفاق نمیوفتد.

برای مقادیر 0.5 و 0.1 در یک خانه ثابت می‌ماند، زیرا در این مقادیر ارزش پیشنهادی انقدر کم است که بی حرکت ماندن را بهتر از حرکت به سمت هدف میبیند، ولی در مقدار 0.9 چون هنوز ارزش حرکت رو به آینده بیشتر است، به سمت هدف حرکت میکند.

سوال ۳

با اجرای الگوریتم policy iteration با پارامترهای های گفته شده در custom_map_3:

soal 3.txt

با فایل بالا، زمانی که True است به معنای غیر قطعی بودن حرکات است.(زمین سر است.) و با بودن این شرایط، تابع ارزش حالت داری حرکات رندوم بیشتری است و رفته رفته V[state] ما به حالت stable تری میرسد. و در نقاط hole ارزش تابع منفی است. ولی در False حرکت به سمت reward به صورت قطعی انجام میشود.

سوال ۴

با اجرای الگوریتم policy iteration با پارامترهای های گفته شده در custom_map_4:

soal 4.txt

با فایل بالا، زمانی که True است به معنای غیر قطعی بودن حرکات است.(زمین سر است.) و با بودن این شرایط، تابع ارزش حالت داری حرکات رندوم بیشتری است و رفته رفته V[state] ما به حالت stable تری میرسد. و در نقاط hole ارزش تابع منفی است. ولی در False حرکت به سمت reward به صورت قطعی انجام میشود.

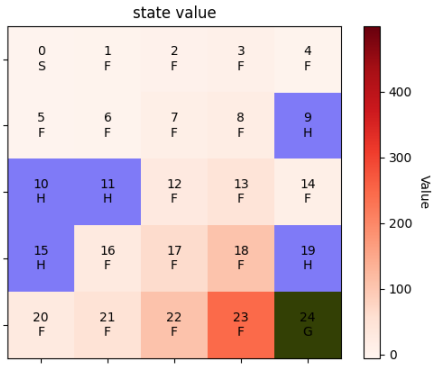
سوال ۵ و ۶

با اجرای الگوریتم policy iteration با پارامترهای های گفته شده در custom_map_5 و custom_map_6:

با توجه به داده ها ارزش تابع حالت، ممکن است تغییر کند اما از نظر نسبی تغییر نمیکند و یک مسیر معین به سمت هدف تعیین میشود.

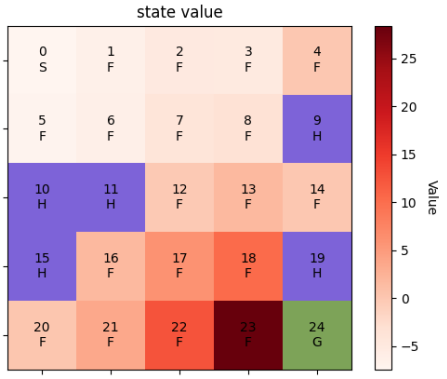
سوال ۷

با اجرای الگوریتم policy iteration با پارامترهای های گفته شده در custom_map_7:

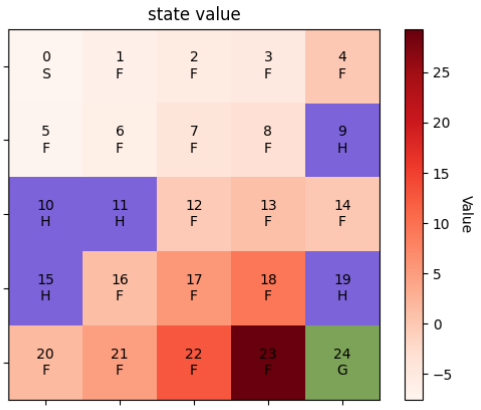


با توجه به عکس بالا، زمانی که policy iteration را اجرا میکنیم، state value ها در نقطه 12 ممکن است به سمت پایین و سمت چپ نیز حرکت کند، به همین دلیل در بعضی از مواقع به hole شماره 15 وارد شود.

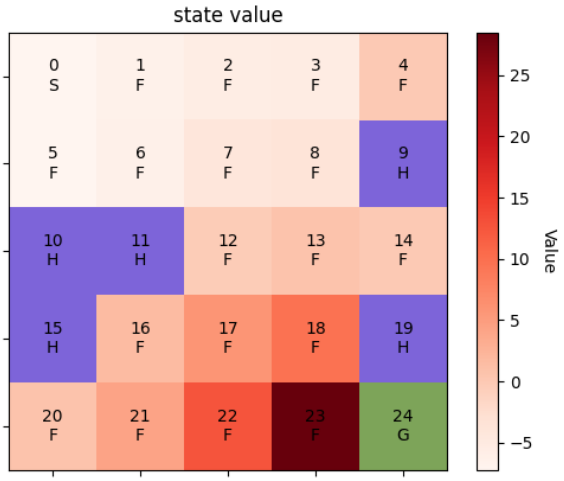
first 500



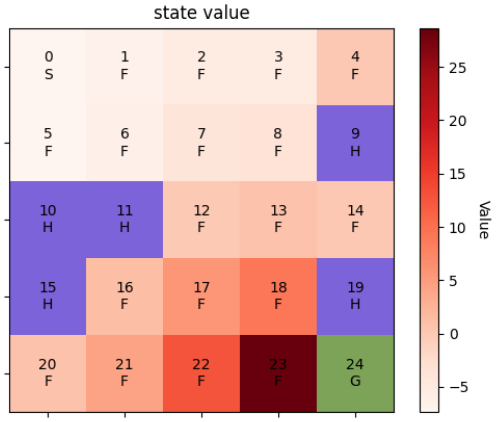
every 500



First 5000



Every 5000

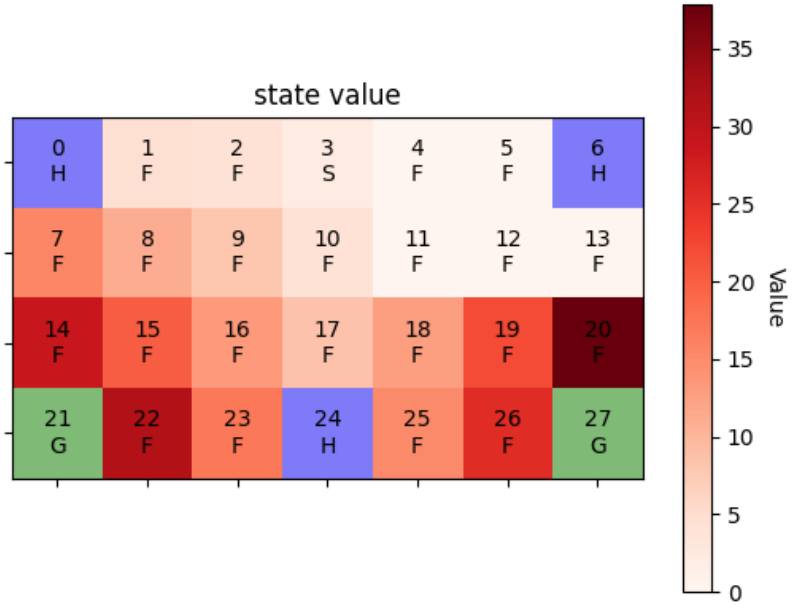


در مقادیر تغییر داده شده در num_episodes به صورت تقریبی برابر بوده و رنگ به مقدار کمی متفاوت هستند.

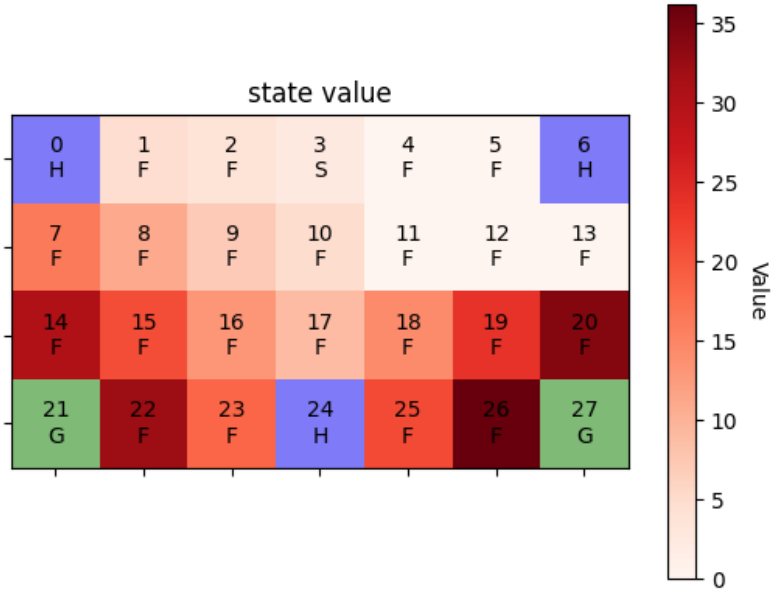
سوال ۸

با اجرای الگوریتم policy iteration با پارامترهای های گفته شده در custom_map_8:

First 1000



Every 1000



همانطور که در عکس های بالا مشخص است در first راه ها را پس از یک بار پیمایش ارزش گذاری میکند و این در جایی معلوم میشود که خانه 26 که کنار هدف است در first کمرنگ تر است و در every به دلیل محاسبه پیاپی با ارزش بیشتری است.