

بررسی پیمایش ارزش و سیاست در محیط دریاچه یخ‌زده

حامد محمدزاده

دانشگاه فردوسی مشهد

Hamedrq7@gmail.com

چکیده

کلمات کلیدی:

1. مقدمه

در مسئله Frozen Lake محیط grid بندی شده است و عامل که در ابتدای بازی در خانه شماره صفر که اولین خانه از بالا سمت چپ است، سعی دارد در هر مرحله از زمان، به یکی از خانه‌های بالا، پایین، راست و چپ برود به طوری که نهایتاً به خانه شماره پانزده، خانه پایین سمت راست برسد. محیط بازی به شانزده خانه به طور چهار در چهار تقسیم شده که بعضی از آن‌ها خانه عادی، بعضی سوراخ و خانه شماره پانزده هم جایزه عامل که می‌خواهد به آن برسد قرار دارد. نقشه بازی در شکل شماره ۱ موجود است، اگر عامل به جهتی برود که خانه‌ای آنجا موجود نباشد، در موقعیت خود باقی می‌ماند، همچنین اگر عامل به خانه سوراخ برسد، بازی مجدداً با شرایط اولیه تکراری آغاز می‌شود و اگر عامل به خانه هدف، خانه شماره پانزده برسد، پاداش ۱ دریافت کرده و باز مجدداً آغاز می‌شود. همچنین تمام حرکات دیگر که به خانه پانزده ختم نشوند امتیازی برای عامل در بر ندارند.

در حالت دیگری از این مسئله که محیط غیر قطعی است، عامل به دلیل لغزندگی همیشه در جهت انتخاب شده حرکت نمی‌کند و به عنوان مثال با انتخاب حرکت به راست، با احتمال ۰.۳۳ به ممکن است راست، بالا یا پایین برود.

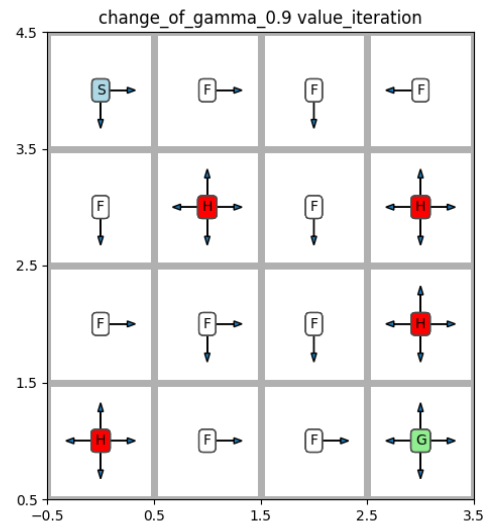
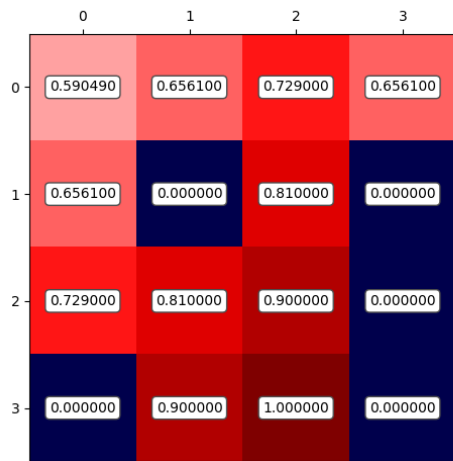
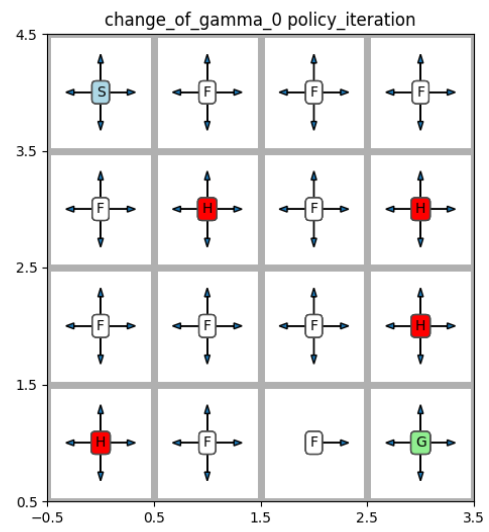
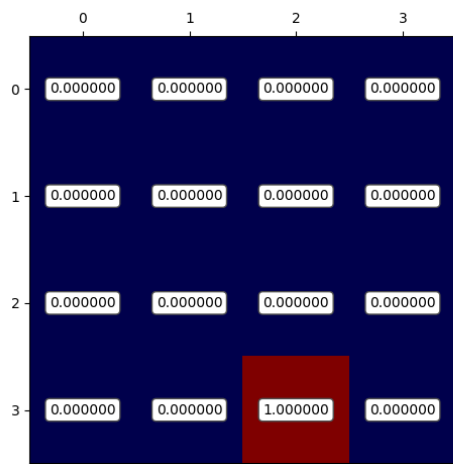


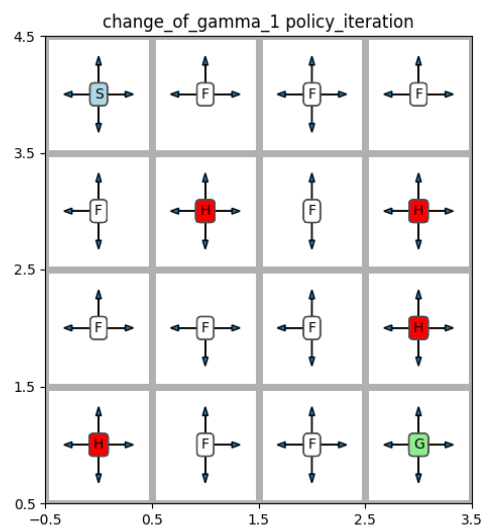
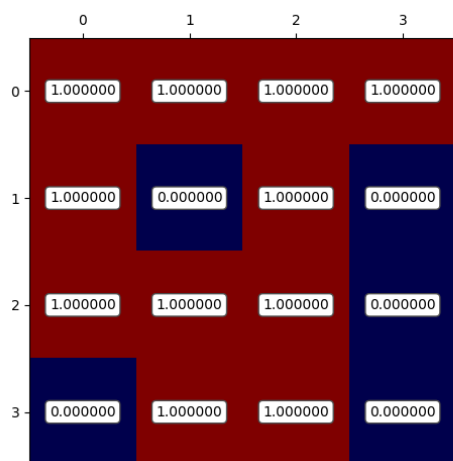
۲. متن یا بدنه اصلی مقالات

در ادامه به بررسی میپردازیم:

2-1. تغییر مقدار ضریب تنزیل

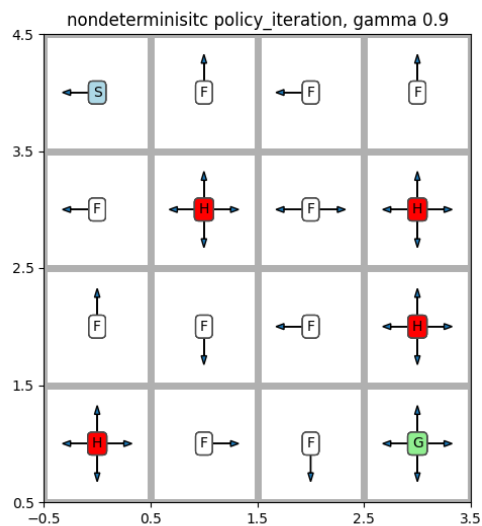
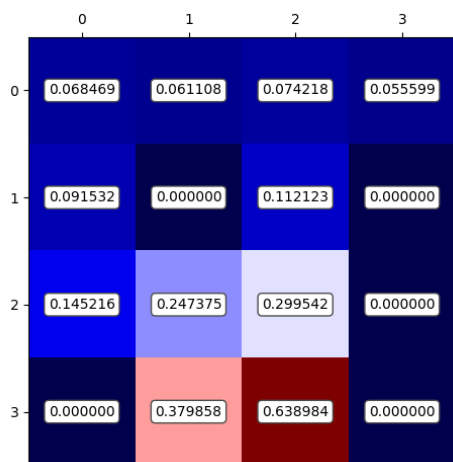
الگوریتم	ضریب تنزیل	شماره مرحله الگوریتم	میانگین بازگشت	میانگین حرکات عامل
value iteration	0	1	0.039	7.7
value iteration	0	2	0.023	7.342
value iteration	0.1	1	0.08	7.658
value iteration	0.1	2	0.213	7.401
value iteration	0.9	1	0.096	7.611
value iteration	0.9	2	0.226	7.618
value iteration	1	1	0.089	7.746
value iteration	1	2	0.206	8.83
policy iteration	0	1	0.035	7.33
policy iteration	0	2	-	-
policy iteration	0.1	1	1	6.973
policy iteration	0.1	2	1	6
policy iteration	0.9	1	1	6
policy iteration	0.9	2	1	6
policy iteration	1	1	1	6





2-2. حل مسئله به صورت غیر قطعی

سیاست بهینه بدست آمده در شکل و رسم شده.



2-3. جریمه منفی برای هر تغییر موقعیت عامل

```
-----
PI - iteration 1 gamma 0.9
avg_return: -0.29999999999999999
avg_steps: 6.0
-----
PI - iteration 2 gamma 0.9
avg_return: -0.29999999999999999
avg_steps: 6.0
out of loop
-----
VI - iteration 1 gamma 0.9
avg_return: -0.2591
avg_steps: 7.246
-----
VI - iteration 2 gamma 0.9
avg_return: -0.26655
avg_steps: 7.418
-----
VI - iteration 3 gamma 0.9
avg_return: -0.27664999999999995
avg_steps: 7.08
-----
VI - iteration 4 gamma 0.9
avg_return: -0.32469999999999993
avg_steps: 6.994
-----
VI - iteration 5 gamma 0.9
avg_return: -0.29999999999999999
avg_steps: 6.0
-----
VI - iteration 6 gamma 0.9
avg_return: -0.29999999999999999
avg_steps: 6.0
-----
VI - iteration 7 gamma 0.9
avg_return: -0.29999999999999999
avg_steps: 6.0
```

2-4. جریمه منفی برای قرارگیری در سوراخ:

```

-----
PI - iteration 1  gamma  0.9
avg_return:  1.0
avg_steps:   6.0
-----
PI - iteration 2  gamma  0.9
avg_return:  1.0
avg_steps:   6.0
out of loop
1.0 6.0
-----
VI - iteration 1  gamma  0.9
avg_return:  -1.703
avg_steps:   7.695
-----
VI - iteration 2  gamma  0.9
avg_return:  -1.385
avg_steps:   7.428
-----
VI - iteration 3  gamma  0.9
avg_return:  -0.476
avg_steps:   6.954
-----
VI - iteration 4  gamma  0.9
avg_return:  1.0
avg_steps:   6.988
-----
VI - iteration 5  gamma  0.9
avg_return:  1.0
avg_steps:   6.0
-----
VI - iteration 6  gamma  0.9
avg_return:  1.0
avg_steps:   6.0
-----
VI - iteration 7  gamma  0.9
avg_return:  1.0
avg_steps:   6.0

```