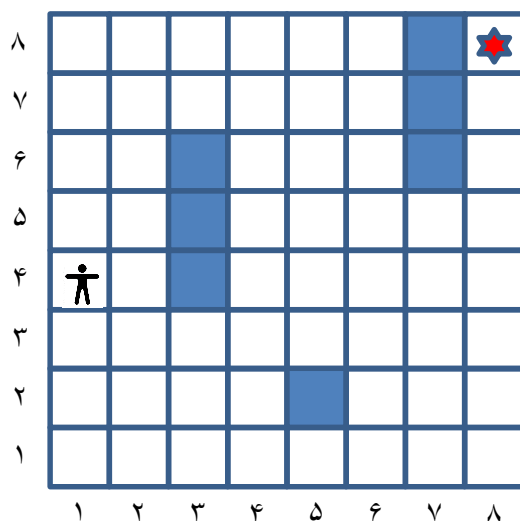


یادگیری تقویتی - نیمسال دوم ۱۴۰۱-۱۴۰۰
تکلیف شماره ۱ - تحویل یکشنبه ۱۴۰۰/۱۲/۲۲

آشنائی با برنامه نویسی پویا در یادگیری تقویتی

یک دنیای شبکه ای ۸ در ۸ همانند شکل ۱ در نظر بگیرید. عاملی وجود دارد که در هر یک از سلولهای این شبکه می



شکل ۱

تواند قرار داشته باشد. به عنوان مثال در شکل ۱ عامل در سلول (۱و۴) می باشد. مقصد نیز در سلول (۸و۸) قرار دارد. برخی از سلولهای این دنیا دارای مانع هستند که در شکل با سلولهای توپر نشان داده شده اند. عامل می تواند به راست، چپ، بالا و پائین حرکت کند و در صورتی که به مرز این دنیا و یا با یک مانع برخورد کند در همان سلولی که بوده باقی می ماند. پاداش انجام هر حرکت نیز صفر و پاداش رسیدن به هدف ۱۰۰ می باشد.

۱- با استفاده از الگوریتم ارزیابی سیاست، سیاستی که انجام هر یک از حرکات احتمال یکسان دارد را ارزیابی نموده و مقادیر آن را نمایش دهید. ضریب تخفیف را ۰٫۹، نظر بگیرید.

۲- با استفاده از الگوریتم تکرار سیاست با ضریب تخفیف ۰٫۹، سیاست بهینه و مقادیر تابع مقدار-حالت بهینه را یافته و نمایش دهید. آیا برای یافتن سیاست بهینه لازم شده است که تابع مقدار-حالت بهینه را بیابید؟ (سعی کنید سیاست را بصورت گرافیکی نمایش دهید، بطور مثال با علائم $<$, \wedge , $>$, \vee).

۳- با استفاده از الگوریتم تکرار مقدار با ضرب تخفیف ۰،۹ سیاست بهینه را بیابید. آیا از لحاظ اجرا تفاوتی بین این الگوریتم و الگوریتم تکرار سیاست مشاهده کرده اید؟

۵- تمرین ۳-۱۲ کتاب را انجام دهید؟

برای پیاده سازیها از زبان پایتون استفاده نمائید. مقادیر اعشاری را تا سه رقم اعشار بیشتر نمایش ندهید. (همیشه برنامه های خود را از کوچک به بزرگ رشد دهید و هنگامی که از صحت یک قسمت مطمئن شدید بخش دیگری را به آن اضافه نمائید).

پاسخها به سؤالات فوق و نتایج آزمایشهای خود را در گزارشی به فرمت PDF به همراه برنامه های خود تحویل نمائید. یک فیلم کوتاه از اجرای برنامه توسط خودتان به همراه توضیح آزمایشهایی که انجام داده اید ارسال نمائید.

لطفاً به نکات زیر توجه نمائید:

- تکالیف تا دو روز بعد از موعد مقرر قابل تحویل می باشند و به ازاء هر روز تأخیر ۱۰٪ از نمره آن کسر می گردد.
- از نوشتار محاوره ای و دارای غلط املائی پرهیز نمائید چون باعث کسر نمره خواهد شد.
- از درج عکس دستنوشته در گزارش خودداری نمائید، هیچ نمره ای به آن تعلق نخواهد گرفت.
- فقط به تکالیفی که در سامانه تحویل داده شوند نمره داده خواهد شد. تحت هیچ شرایطی تکلیف خود را ایمیل نکنید.
- به تکالیف مشابه و دانلود شده نمره ای تعلق نخواهد گرفت.

موفق باشید

پالهنگ