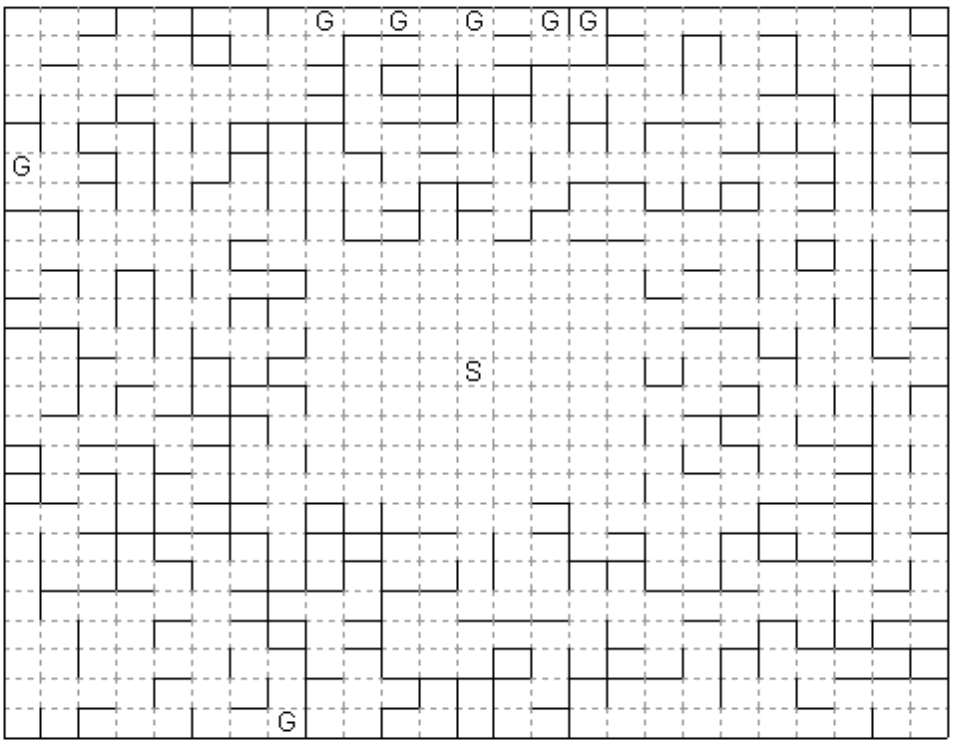
وزارت علوم ، تحقیقات و فناوری
دانشگاه تحصیلات تکمیلی علوم پایه
گاوازنگ - زنجان

به نام او

تکلیف سوم- طراحی سیستم‌های هوشمند

دانشگاه تحصیلات تکمیلی علوم پایه زنجان-بهار ۹۶

فرض کنید که شخصی در محیط پر پیچ و خمی (maze) قرار دارد و هدف این است که عامل بتواند از نقطه شروع به هدف برسد. به طور مثال مسئله maze با اندازه‌ی ۲۵×۲۵ را در شکل ۱ در نظر بگیرید. مکان اولیه عامل با استفاده از برچسب S نمایش داده شده است و مکان‌های هدف با استفاده از G نمایش داده شده است (رسیدن به یک هدف کافی است). عامل قادر است که در هر چهار جهت شمال، شرق، جنوب و غرب حرکت نماید.



شکل ۱ - مسئله maze با اندازه‌ی ۲۵×۲۵

ورودی برنامه به صورت زیر است:

```
5 5
0 1 1 0
1 1 1 0
1 0 1 0
1 1 1 0
1 1 0 0
0 1 0 1
0 1 1 1
```

```
1 1 1 0
1 1 1 1
1 0 0 1
0 1 1 1
1 1 0 1
0 0 0 1
0 0 0 1
0 1 0 0
0 0 0 1
0 1 1 1
1 0 0 0
0 0 1 0
1 0 0 1
0 0 1 0
1 0 1 1
1 0 1 0
1 0 1 0
1 0 0 0
13 25
```

که در آن خط اول بیانگر تعداد سطر و ستون‌های جدول maze است. خطوط ۲ تا ۲۵ بیانگر این است که هر سلول maze از کدام یک از جهات توسط مانع پوشیده شده (که با صفر نمایش داده می‌شود) و یا پوشیده نشده است (با یک نمایش داده می‌شود). در خط آخر، عدد اول بیانگر مکان اولیه عامل است و عدد(های) بعدی بیانگر مکان(های) هدف است. سلول‌های جدول maze به ترتیب به صورت ستونی شماره گذاری شده‌اند به گونه‌ای که هر سلول آن متناظر با یک عدد است. مکان اولیه عامل و مکان‌های هدف با استفاده از شماره سلول متناظر در ورودی نمایش داده می‌شود. نمونه maze متناظر با ورودی بالا در شکل ۲ نشان داده شده است.

موارد زیر را برای دو مسئله maze ۵×۵ و ۲۵×۲۵ انجام دهید. مسئله‌های مذکور به صورت فایل متنی همراه با فایل تکلیف قرار خواهد گرفت.

1	6	11	16	21
2	7	12	17	22
3	8	S 13	18	23
4	9	14	19	24
5	10	15	20	G 25

شکل ۲ - مسئله maze با اندازه‌ی ۵×۵

الف) با فرض اینکه عامل محیط را کامل می‌شناسد، سیاست بهینه در هر قدم برای عامل را با استفاده از الگوریتم Value Iteration به دست آورید. پس از به دست آوردن سیاست بهینه برای تمام خانه‌های دنیای maze، عامل را در نقطه شروع قرار داده و اجازه دهید با استفاده از سیاست بهینه آموزش دیده حرکت نماید. مسیر عامل از نقطه شروع تا هدف مسیر بهینه می‌باشد؟ چرا؟

ب) با فرض اینکه عامل محیط را کامل می‌شناسد، سیاست بهینه در هر قدم برای عامل را با استفاده از الگوریتم Policy Iteration به دست آورید. پس از به دست آوردن سیاست بهینه برای تمام خانه‌های دنیای maze، عامل را در نقطه شروع قرار داده و اجازه دهید با استفاده از سیاست بهینه آموزش دیده حرکت نماید. مسیر عامل از نقطه شروع تا هدف مسیر بهینه می‌باشد؟ چرا؟

ج) پاسخ‌های حاصل از الف و ب را با هم مقایسه نمایید. سپس تحلیل نمایید که کدامیک از این روش‌ها بر دیگری برتری دارد و چرا؟

چ) فرض کنید که عامل محیط را نمی‌شناسد و تنها قادر است خانه‌ای که در آن قرار گرفته است را تشخیص دهد، سیاست بهینه برای عامل را با استفاده از الگوریتم Q بیابید. دقت کنید که در این صورت از ورودی برنامه تنها برای شبیه‌سازی پاسخ محیط به ازای هر عمل عامل استفاده نمائید. پس از به دست آوردن سیاست بهینه با استفاده از الگوریتم Q، عامل را در نقطه شروع قرار داده و اجازه دهید با استفاده از سیاست بهینه آموزش دیده حرکت نماید. مسیر عامل از نقطه شروع تا هدف مسیر بهینه می‌باشد؟ چرا؟

د) برای انتخاب عمل در هر قدم در طی الگوریتم Q از چه روشی استفاده می‌نمایید و چرا؟