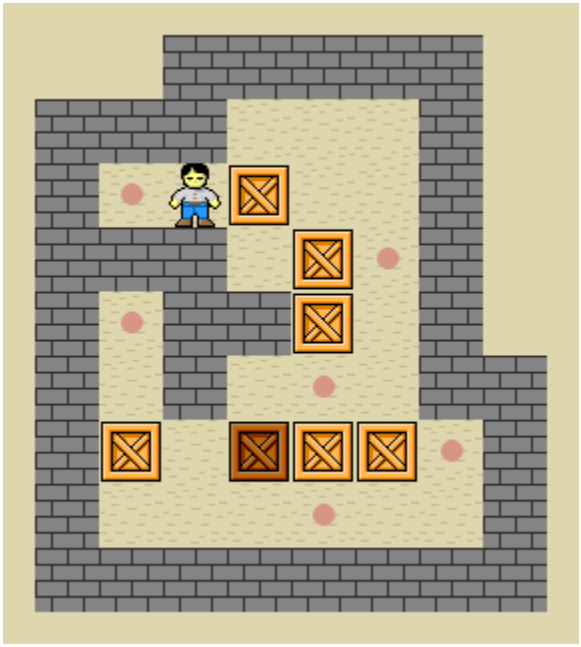
وزارت علوم ، تحقیقات و فناوری
دانشگاه تحصیلات تکمیلی علوم پایه
گاوازنگ – زنجان

به نام او

تکلیف اول

طراحی سیستم‌های هوشمند

پازل sokoban سال ۱۹۸۱ توسط ایمابایاشی معرفی شد. این پازل شامل یک محیط مستطیلی است که در هر خانه آن یا دیوار قراردارد و یا خانه باز می‌باشد. همچنین محیط شامل عامل و چندین جعبه نیز می‌باشد که جایگاه آن‌ها در محیط مشخص می‌باشد. هدف اصلی در این پازل این است که عامل بتواند از نقطه شروعی که در آن قرار گرفته است با حرکت دادن جعبه ها آن‌ها را به مکان نهایی جعبه ها برساند. مکان نهایی جعبه ها نیز در محیط مشخص می‌باشد (شکل ۱).



شکل ۱ - محیط پازل sokoban

ورودی مسئله به شرح زیر است: خط اول شامل دو عدد است که بیانگر تعداد سطرها (n) و ستون های (m) محیط است. در n سطر بعدی، m کاراکتر وجود دارد که هر یک از این کاراکترها می تواند یکی از موارد زیر باشد که به شرح زیر تفصیل می گردد:

- #: بیانگر دیوار است
- @: محل قرار گیری جعبه ها
- .: مربع هایی که باز هستند و عامل می تواند از طریق آن‌ها حرکت نماید
- S: محل قرارگیری اولیه عامل

- X: مقصد نهایی جعبه ها

به طور مثال یک نمونه از ورودی مسئله به شرح زیر است:

```
5 12
############
######..X###
#S....@...##
######..####
############
```

عامل در محیط در صورتی که مربع کناری خالی باشد به سمت بالا، پایین، چپ و یا راست حرکت می نماید. همچنین عامل قادر است جعبه را در چهار جهت بالا، پایین، چپ و یا راست در صورتی خالی بودن مربع تکان دهد.

در صورتی که عامل پس از حرکت با دیوار برخورد نماید امتیاز ۱۰-، برخورد با جعبه ۱۰+، و رسیدن به مقصد نهایی جعبه همراه با جعبه ۱۰۰+ و در غیر اینصورت امتیاز ۰ می گیرد. دقت کنید که در صورتی که عامل بدون جعبه به مقصد نهایی برسد امتیازی دریافت نمی کند. همچنین در صورتی که عامل همراه با جعبه وارد مقصد نهایی شود وارد حالت absorb state شده است.

در این تمرین ورودی ها تنها شامل یک جعبه و یک مکان هدف است. دانشجویانی که بتوانند محیط های شامل چند جعبه هدف و چند مکان هدف را نیز حل نمایند، نمره اضافی دریافت می‌کنند. موارد زیر را برای مسئله هایی که به صورت فایل متنی همراه با فایل تکلیف خواهد گرفت، انجام دهید و نتایج را گزارش دهید.

الف) فرض کنید که نقشه کل محیط را از قبل می دانید (به عبارتی دیگر ورودی مسئله را می توانید در شروع کار استفاده نمایید) و با استفاده از الگوریتم value iteration و policy iteration سیاست بهینه را برای عامل پیدا کنید. پس از یافتن سیاست بهینه، اجازه دهید عامل با سیاست بهینه در محیط حرکت نماید و نتایج حاصل را گزارش داده و تحلیل نمایید.

ب) فرض کنید که نقشه کل محیط را از قبل نمی دانید، به عبارتی دیگر از ورودی مسئله و تابع reward تنها برای گزارش reward لحظه ای و نتیجه فعالیت در لحظه استفاده کنید. با استفاده از الگوریتم Q-learning سیاست بهینه را برای عامل پیدا کنید. پس از یافتن سیاست بهینه، اجازه دهید عامل با سیاست بهینه در محیط حرکت نماید و نتایج حاصل را گزارش داده و تحلیل نمایید..

دانشجویان محترم، هر یک از افراد نویسنده‌ی پروژه، می بایست اشراف کامل و جامع بر پیاده‌سازی کار داشته باشند. همچنین به پروژه هایی که شامل کپی از هر منبعی باشد نمره صفر تعلق خواهد گرفت.