

**دانشکده: مهندسی کامپیوتر**

**موضوع: داک تمرین عملی پنجم AI**

**علی شکوهی**

**شماره دانشجویی: 40052147**

**متغیرها:**

**Qvalues: یک Counter برای نگهداری مقادیر Q برای هر حالت.**

**landa: تاثیر گذاری فراموشی در به‌روزرسانی Q-value.**

**epsilon: برای استفاده در روش ε-greedy برای انتخاب عمل.**

**alpha: نرخ یادگیری در به‌روزرسانی Q-value.**

**states: مجموعه حالات شناخته‌شده تاکنون.**

**iterations: تعداد اجراهای الگوریتم RL.**

**bins و discrete\_states: برای تبدیل حالت‌های پیوسته به حالت‌های گسسته.**

**actions\_number: تعداد عمل‌ها (در اینجا ۲).**

**Qtable: جدول Q-values برای هر حالت و عمل.**

**توابع پیاده‌سازی شده:**

**policy(state): پیاده‌سازی روش ε-greedy برای انتخاب عمل.**

**get\_all\_actions: بازگرداندن تمام عمل‌های ممکن.**

**convert\_continuous\_to\_discrete(state): تبدیل حالت پیوسته به حالت گسسته.**

**compute\_reward(prev\_info, new\_info, done, observation): محاسبه پاداش بر اساس اطلاعات مشاهده‌شده.**

**get\_action(state): دریافت عمل بر اساس حالت فعلی.**

**maxQ(state): برگرداندن بزرگترین مقدار Q برای یک حالت.**

**max\_arg(state): برگرداندن عملی که مقدار Q بیشینه را دارد.**

**update(reward, state, action, next\_state): به‌روزرسانی جدول Q-values.**

**update\_epsilon\_alpha: به‌روزرسانی مقادیر epsilon و alpha.**

**run\_with\_policy(landa): اجرای الگوریتم RL با روش Policy.**

**run\_with\_no\_policy(landa): اجرای بازی بدون تاثیر Policy.**

**run: اجرای الگوریتم RL و بازی بدون تاثیر Policy.**

**ممکن است عملکرد الگوریتم در 3000 ایتریشن نتایج بهینه را نداشته باشد.**

**بهبود عملکرد با تنظیم مقادیر پارامترها ممکن است انجام شود.**