**DESCRIZIONE PROGETTO**

**NOME PROGETTO**: ROBOT WAREHOUSE – AI

**SCENARIO DI UTILIZZO**: Una società di e-commerce sta costruendo un nuovo magazzino e la società vorrebbe che tutte le operazioni di prelievo nel nuovo magazzino fossero eseguite da robot di magazzino. L’operazione di "prelievo" consiste nel raccogliere singoli articoli da varie posizioni del magazzino al fine di soddisfare gli ordini dei clienti. Dopo aver prelevato gli articoli dagli scaffali, i robot devono portare gli articoli in una posizione specifica all'interno del magazzino, utilizzando il percorso più breve possibile (per garantire la massima efficienza e produttività), dove gli articoli possono essere imballati per la spedizione.

**TECNICA AI**: Q-learning (paradigma di machine learning noto per l’apprendimento tramite rinforzo)

**IMPLEMENTAZIONE TECNICA**

**Libreria richiesta**

“numpy”

**Definizione dell'ambiente**

L'ambiente è costituito da stati, azioni e ricompense. Gli stati e le azioni sono input per l'agente AI di Q-learning, mentre le azioni possibili sono gli output dell'agente AI. Gli stati nell'ambiente sono tutte le possibili posizioni all'interno del magazzino. Alcune di queste posizioni servono per riporre gli oggetti (quadrati neri), mentre altre sono corridoi che il robot può utilizzare per viaggiare in tutto il magazzino (quadrati bianchi). Il quadrato verde indica l'area di imballaggio e spedizione dell'articolo. I quadrati nero e verde sono stati terminali!

Immagine che contiene testo, frutta

Descrizione generata automaticamente

**L'obiettivo dell'agente AI**

È apprendere il percorso più breve tra l'area di confezionamento degli articoli e tutte le altre posizioni del magazzino in cui il robot può viaggiare. Come mostrato nell'immagine sopra, ci sono 121 possibili stati (posizioni) all'interno del magazzino. Questi stati sono disposti in una griglia contenente 11 righe e 11 colonne. Ogni posizione può quindi essere identificata dal suo indice di riga e colonna.

**Azioni**

Le azioni disponibili per l'agente AI consistono nel muovere il robot in una delle quattro direzioni:

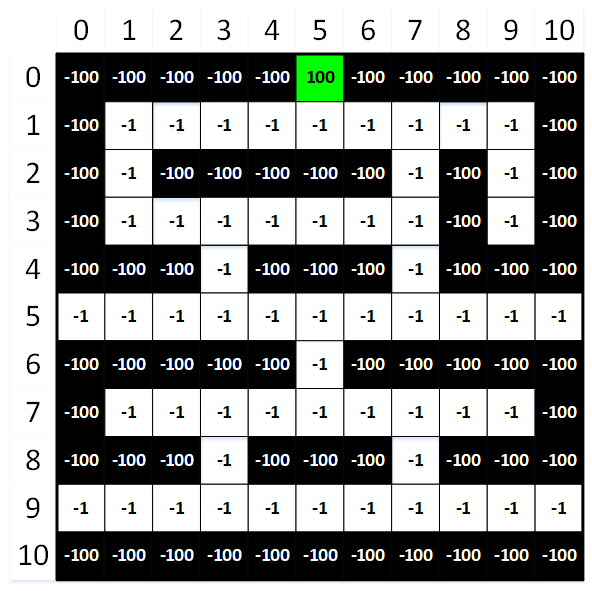
[Su, Destra, Giù, Sinistra]

Ovviamente, l'agente AI deve imparare a evitare di guidare nelle posizioni di stoccaggio degli oggetti (ad esempio, scaffali).

**Ricompense**

L'ultima componente dell'ambiente che dobbiamo definire sono le ricompense. Per aiutare l'agente AI ad apprendere, a ogni stato (posizione) nel magazzino viene assegnato un valore di ricompensa. L'agente può iniziare da qualsiasi quadrato bianco, ma il suo obiettivo è sempre lo stesso: massimizzare le sue ricompense totali.

Le ricompense negative (cioè le punizioni) vengono utilizzate per tutti gli stati tranne l'obiettivo. Questo incoraggia l'IA a identificare il percorso più breve verso l'obiettivo riducendo al minimo le sue punizioni.



Per massimizzare i suoi premi cumulativi (riducendo al minimo le sue punizioni cumulative), l'agente AI dovrà trovare i percorsi più brevi tra l'area di imballaggio dell'articolo (quadrato verde) e tutte le altre posizioni nel magazzino dove il robot è autorizzato a viaggiare (quadrati bianchi). L'agente dovrà anche imparare a evitare di schiantarsi contro le posizioni di deposito degli oggetti (quadrati neri).

**Addestramento**

Il nostro prossimo compito è che il nostro agente AI apprenda il suo ambiente implementando un modello di Q-learning. Il processo di apprendimento seguirà questi passaggi:

1. Scegli uno stato casuale, non terminale (quadrato bianco) per l'agente per iniziare questo nuovo episodio.
2. Scegli un'azione (sposta in alto, a destra, in basso o a sinistra) per lo stato corrente. Le azioni verranno scelte utilizzando un algoritmo *epsilon greedy*. Questo algoritmo solitamente sceglierà l'azione più promettente per l'agente AI, ma a volte sceglierà un'opzione meno promettente per incoraggiare l'agente a esplorare l'ambiente.
3. Esegui l'azione scelta e passa allo stato successivo (ovvero passa alla posizione successiva).
4. Ricevi la ricompensa per il passaggio al nuovo stato e calcola la differenza temporale.
5. Aggiorna il valore Q per la coppia di stato e azione precedente.
6. Se il nuovo stato (corrente) è uno stato terminale, vai al n. 1. Altrimenti, vai al n. 2.

L'intero processo verrà ripetuto in 1000 episodi. Ciò fornirà all'agente AI un'opportunità sufficiente per apprendere i percorsi più brevi tra l'area di imballaggio degli articoli e tutte le altre posizioni nel magazzino in cui il robot può viaggiare, evitando allo stesso tempo di schiantarsi in qualsiasi posizione di stoccaggio degli articoli.