

Treasure Trail.



Espacio de acción	discreto(4)
Espacio de observación	Discreto(100)
Importar	<code>gym.make("TreasureTrail-v1")</code>

Descripción.

Existen 3 diamantes en este entorno, ubicados en [43, 75, 89], además de 15 trampas, ubicadas en [2, 12, 24, 40, 42, 48, 54, 55, 56, 57, 60, 63, 84, 86, 90, 92, 98, 99] en un grid world de 10x10. El agente empieza el recorrido en la posición 0. El objetivo del agente es encontrar todos los diamantes que se encuentran en el entorno. Cuando encuentra uno de estos objetos, se le da al agente una recompensa positiva de 1. El episodio finaliza cuando el agente logra encontrar los 3 diamantes en el entorno, o, cuando cae en una trampa.

Espacio de observación.

Hay 100 espacios de observación discretos en este entorno. Siendo las posiciones del laberinto, en las cuales están ubicadas 3 diamantes, y 15 trampas. Cada observación se retorna como una tupla de 4 elementos:

1. Un entero, que representa el siguiente estado al ejecutar cierta acción.
2. Un flotante, que representa la recompensa obtenida al ejecutar dicha acción.
3. Un booleano, que representa si el episodio ha terminado, o no.
4. Un booleano, que representa si el episodio esta truncado, o no.

Nota: Los estados donde se ubican las recompensas (diamantes), cambian durante la ejecución del entrenamiento.

Espacio de Acción:

El espacio de acción está definido en el rango de 0 a 3. Definido de la siguiente manera:

0: Izquierda.

1: Abajo.

2: Derecho.

3: Arriba.

Estado Inicial:

El episodio comienza en el estado 0.

Recompensas:

El agente obtiene recompensas cuando llega a los estados donde están ubicados los diamantes.

Fin del episodio:

El episodio termina cuando ocurre lo siguiente:

1. Si consigue todas las recompensas (3), el episodio termina.
2. Si el agente llega a un estado donde está ubicada una trampa, termina el episodio, y vuelve al estado inicial.

Descripción Del Agente.

El agente utilizado para resolver este entorno fue el agente Montecarlo no determinista. Se seleccionó este agente por su eficiencia en la exploración ya que los algoritmos de Monte Carlo son conocidos por su capacidad de exploración en entornos desconocidos. Dado que las recompensas están dispersas en el entorno, un agente de Monte Carlo podría ser efectivo en encontrar y explorar todas las áreas del entorno en busca de estas recompensas. También tomamos en cuenta su capacidad de generalización, puesto que dependiendo del entorno y la forma en que se definen las recompensas dispersas, un agente de Monte Carlo puede aprender a generalizar la mejor política de acción para maximizar la recompensa total en situaciones similares pero no idénticas. Esto puede ser beneficioso cuando las recompensas dispersas se encuentran en diferentes ubicaciones o tienen diferentes condiciones, pero comparten ciertos patrones o características comunes.

Para complementar cabe mencionar que los métodos de Monte Carlo trabajan con episodios completos, lo que significa que pueden aprender a partir de la experiencia acumulada en múltiples episodios. Si el entorno con recompensas dispersas requiere una secuencia de acciones específicas para alcanzar y obtener las recompensas, un agente de Monte Carlo podría aprender a través de múltiples episodios cómo optimizar su política de acción para maximizar la recompensa total a largo plazo

Durante la demostración del entorno, como ya se comentó anteriormente se utilizó el agente Montecarlo no determinista, que se caracteriza por ser de tipo tabular. Para lograr resolver el entorno el agente requiere aproximadamente entre doscientos mil y trescientos mil episodios. Sin embargo, es importante destacar que este número puede variar dependiendo de las características específicas del entorno y del agente utilizado.

El código fuente del entorno, así como el agente usado para resolverlo se encuentra en: <https://github.com/NatanaelRojo/computer-system-project.git>

Un video de una pequeña demo del entorno se puede ver en:

<https://youtu.be/ZW1Yj0C33SQ>

Natanael Rojo C.I: 26.488.388

Heberto Gutiérrez C.I: 24.752.816