

Instrucciones:

- Esta es una actividad en grupos de 3 personas máximo
- No se permitirá ni se aceptará cualquier indicio de copia. De presentarse, se procederá según el reglamento correspondiente.
- Tendrán hasta el día indicado en Canvas.

Task 1

Responda a cada de las siguientes preguntas de forma clara y lo más completamente posible.

1. ¿Qué es Prioritized sweeping para ambientes determinísticos?
2. ¿Qué es Trajectory Sampling?
3. ¿Qué es Upper Confidence Bounds para Árboles (UCT por sus siglas en inglés)?

Task 2

En este laboratorio, compararán el rendimiento de Dyna-Q+ y MCTS, dos de los algoritmos que vimos en clase, utilizando el entorno de FrozenLake-v1 de la biblioteca Gymnasium. Analizará y graficará las recompensas por episodio y responderá las preguntas que aparecen al final para asegurar su comprensión de los algoritmos.

Instrucciones

1. Implementación de MCTS:
 - a. Implemente un algoritmo de búsqueda de árbol de Monte Carlo (MCTS) para resolver el entorno FrozenLake-v1.
 - b. Use una estructura de árbol para simular diferentes secuencias de acciones a partir del estado actual.
 - c. Para cada secuencia de acciones simulada, implemente una política en un estado terminal, acumule recompensas y propague estas recompensas a través del árbol.
 - d. Seleccione acciones en función de las rutas más prometedoras descubiertas durante la búsqueda.
 - e. Considere usar límites de confianza superior para árboles (UCT) para equilibrar la exploración y la explotación en su búsqueda.
 - f. Implementación de MCTS:
 - g. Implemente un algoritmo de búsqueda de árbol de Monte Carlo (MCTS) para resolver el entorno FrozenLake-v1.
 - h. Use una estructura de árbol para simular diferentes secuencias de acciones a partir del estado actual.
 - i. Para cada secuencia de acciones simulada, implemente una política en un estado terminal, acumule recompensas y propague estas recompensas a través del árbol.
 - j. Seleccione acciones en función de las rutas más prometedoras descubiertas durante la búsqueda.
 - k. Considere usar límites de confianza superior para árboles (UCT) para equilibrar la exploración y la explotación en su búsqueda.
 - l. **Consideración especial:** FrozenLake-v1 tiene dinámica estocástica, lo que significa que las transiciones son probabilísticas. Asegúrese de que su implementación de MCTS maneje estas transiciones probabilísticas de manera adecuada.
2. Implementación de Dyna-Q+:
 - a. Implemente el algoritmo Dyna-Q+ para resolver el entorno FrozenLake-v1.
 - b. Use un enfoque de Q-learning para actualizaciones de valores basadas en experiencias reales.
 - c. Aprenda un modelo del entorno almacenando transiciones y recompensas para pares de estado-acción.

- d. Use el modelo aprendido para generar experiencias simuladas (pasos de planificación) y actualice los valores Q en función de estas simulaciones.
- e. Incorpore una bonificación de exploración en sus valores Q para fomentar la exploración de pares de estado-acción menos visitados.
- f. **Consideración especial:** Ajuste la cantidad de pasos de planificación (parámetro n) y la bonificación de exploración para ver cómo afectan el rendimiento del aprendizaje en un entorno estocástico.
3. Ejecución de experimentos:
 - a. Ejecute varios episodios de FrozenLake-v1 con MCTS y Dyna-Q+ para recopilar datos sobre su rendimiento.
 - b. Registre métricas como la cantidad de episodios exitosos (que alcanzaron el objetivo), las recompensas promedio y la cantidad de pasos dados para alcanzar el objetivo.
4. Análisis gráfico:
 - a. Cree los siguientes gráficos para visualizar y comparar el rendimiento de MCTS y Dyna-Q+:
 - i. **Tasa de éxito en los episodios:** trace el porcentaje de episodios exitosos (que alcanzan la meta) para cada algoritmo en varios episodios.
 - ii. **Recompensa promedio por episodio:** trace la recompensa promedio obtenida por episodio a lo largo del tiempo para ambos algoritmos.
 - iii. **Tasa de convergencia:** trace la cantidad de pasos que se toman para alcanzar la meta (si se logra el éxito) como una función del número de episodios.
 - iv. **Exploración vs. Explotación:** Para Dyna-Q+, trace cómo la bonificación de exploración influye en la política a lo largo del tiempo, mostrando la cantidad de pares de estado-acción visitados vs. la cantidad total de pares de estado-acción.
5. Análisis:
 - a. Compare los resultados de MCTS y Dyna-Q+.
 - b. Analice las fortalezas y debilidades de cada enfoque en el contexto de FrozenLake-v1.
 - c. Considere el impacto de la naturaleza estocástica del entorno en el rendimiento de ambos algoritmos.

Preguntas

1. Estrategias de exploración:
 - a. ¿Cómo influye la bonificación de exploración en Dyna-Q+ en la política en comparación con el equilibrio de exploración-explotación en MCTS? ¿Qué enfoque conduce a una convergencia más rápida en el entorno FrozenLake-v1?
2. Rendimiento del algoritmo:
 - a. ¿Qué algoritmo, MCTS o Dyna-Q+, tuvo un mejor rendimiento en términos de tasa de éxito y recompensa promedio en el entorno FrozenLake-v1? Analice por qué uno podría superar al otro dada la naturaleza estocástica del entorno.
3. Impacto de las transiciones estocásticas:
 - a. ¿Cómo afectan las transiciones probabilísticas en FrozenLake-v1 al proceso de planificación en MCTS en comparación con Dyna-Q+? ¿Qué algoritmo es más robusto a la aleatoriedad introducida por el entorno?
4. Sensibilidad de los parámetros:
 - a. En la implementación de Dyna-Q+, ¿cómo afecta el cambio de la cantidad de pasos de planificación n y la bonificación de exploración a la curva de aprendizaje y al rendimiento final? ¿Se necesitarían diferentes configuraciones para una versión determinista del entorno?

Entregas en Canvas

1. Documento PDF con las respuestas a cada task
 - a. Pueden exportar el JN como PDF si trabajan con esto.
2. Código de la implementación del Task 2
 - a. Si trabaja con JN deje evidencia de la última ejecución
 - b. Caso contrario, deje en comentarios el valor resultante
 - c. Debe entregar el PDF y el JN
3. **POR FAVOR, SI USAN REPOSITORIO TAMBIÉN SUBAN LA VERSIÓN PDF A CANVAS**

Evaluación

1. [1.5 pts] Task 1 (0.5 cada pregunta)
2. [3.5 pts] Task 2