**Módulo: Machine Learning**

**Actividad: Comparación de Modelos de Aprendizaje por Refuerzo**

**Nombre: Alan Giovanni Venegas Turrubiartes**

**Docente: Jesús Manuel Vázquez Nicolá**

**Fecha: 08/09/2024**

Técnicas de aprendizaje por refuerzo.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Técnica | Descripción | Ventajas | Desventajas | Ejemplo de Aplicación |
| Actor-Critic | Combina un actor que toma decisiones y un crítico que evalúa acciones. | Balancea exploración y explotación; política continua. | Complejo de implementar. | Robótica, control de movimientos. |
| Proximal Policy Optimization | Optimización de políticas basada en gradientes proximales para evitar grandes saltos. | Más estable que TRPO; eficiente en entrenamiento. | Costoso en términos computacionales. | Simulaciones y videojuegos. |
| Monte Carlo Tree Search (MCTS) | Construcción de un árbol de decisiones mediante simulaciones aleatorias. | Efectivo para problemas de decisiones complejas sin necesidad de modelo. | Ineficiente en grandes espacios de búsqueda. | Juegos como Go y ajedrez. |
| Double Q-Learning | Variante de Q-learning que usa dos redes Q para reducir el sobreestimado de recompensas. | Más estable que Q-learning simple. | Costoso en recursos debido al uso de dos redes Q. | Control de tráfico, sistemas de recomendación. |
| Trust Region Policy Optimization (TRPO) | Optimización de políticas ajustando gradientes en una región de confianza para evitar grandes actualizaciones. | Garantiza actualizaciones estables. | Más costoso y complicado que PPO. | Control de drones y robots. |

. Comparación de Técnicas Avanzadas de Aprendizaje por Refuerzo

A lo largo del estudio de las diferentes técnicas de aprendizaje por refuerzo, he podido observar la diversidad y complejidad que existe en este campo. En particular, técnicas como Actor-Critic, Proximal Policy Optimization (PPO), Monte Carlo Tree Search (MCTS), Double Q-Learning y Trust Region Policy Optimization (TRPO) representan enfoques avanzados que abordan distintos aspectos del aprendizaje y la toma de decisiones. Al profundizar en cada una de estas técnicas, se hace evidente que, aunque todas persiguen el mismo objetivo general de optimizar las políticas de un agente, lo hacen de maneras muy distintas, cada una con sus fortalezas y debilidades. En este ensayo, reflexionaré sobre los hallazgos más significativos derivados de su comparación.

Uno de los principales aprendizajes que he obtenido es que las técnicas de aprendizaje por refuerzo varían mucho en cuanto a su eficiencia y aplicabilidad según el entorno en el que se empleen. Por ejemplo, Actor-Critic, al combinar los enfoques basados en valor y los enfoques basados en política, ofrece una gran flexibilidad para entornos que requieren adaptabilidad. Sin embargo, he podido notar que esta flexibilidad tiene un costo en términos de complejidad. La implementación de dos componentes que interactúan entre sí (actor y crítico) no es sencilla, y esta dualidad añade una capa adicional de dificultad al proceso de optimización.

Por otro lado, PPO y TRPO han sido reveladoras en cuanto a la importancia de la estabilidad en los procesos de aprendizaje. Ambos métodos, aunque similares en algunos aspectos, se enfocan en realizar actualizaciones de la política de manera controlada, evitando cambios drásticos que puedan desestabilizar el aprendizaje del agente. Me sorprendió especialmente la eficiencia de PPO, que, si bien es demandante en cuanto a recursos computacionales, garantiza una mejora continua y suave en la política del agente. En contraste, aunque TRPO ofrece mejores garantías teóricas de estabilidad, su alto costo computacional lo convierte en una opción menos viable en escenarios donde la eficiencia es crucial. Esto me llevó a reflexionar sobre la importancia de balancear los recursos disponibles con la precisión que se busca en el aprendizaje.

En cuanto a MCTS, me pareció fascinante cómo esta técnica se utiliza en problemas de decisiones complejas, como juegos de estrategia. Su capacidad para simular múltiples posibles caminos antes de tomar una decisión refleja una comprensión profunda del entorno en el que opera el agente. Sin embargo, la principal limitación que encontré es que este método no siempre es eficiente cuando el espacio de búsqueda es demasiado amplio. En esos casos, su principal fortaleza –la exploración exhaustiva de opciones– se convierte en su mayor debilidad, ya que demanda un enorme poder de procesamiento.

Finalmente, el análisis de Double Q-Learning me hizo reflexionar sobre la importancia de minimizar los sesgos en el aprendizaje por refuerzo. Este enfoque, que emplea dos redes Q para evitar el sobreestimado de recompensas, resuelve uno de los problemas más críticos del Q-learning clásico. A través de este método, comprendí mejor cómo pequeños ajustes en los algoritmos pueden llevar a mejoras significativas en la estabilidad y precisión del aprendizaje.

Al comparar todas estas técnicas, he llegado a la conclusión de que no existe una solución única que sea adecuada para todos los problemas. Dependiendo del contexto y de los recursos disponibles, una técnica puede ser más ventajosa que otra. Por ejemplo, en aplicaciones donde se necesita estabilidad, PPO es una opción excelente, mientras que MCTS destaca en problemas de toma de decisiones con múltiples ramificaciones. Sin embargo, en escenarios donde se busca minimizar el sesgo en el aprendizaje, Double Q-Learning podría ser la opción más adecuada.

En resumen, la reflexión sobre estas técnicas me ha permitido apreciar la complejidad y diversidad de soluciones que ofrece el campo del aprendizaje por refuerzo. Cada técnica aborda problemas específicos y tiene aplicaciones en distintos contextos. Como conclusión, el proceso de seleccionar la técnica adecuada no es trivial y requiere un análisis detallado de las necesidades del problema y de los recursos disponibles. La flexibilidad, estabilidad y eficiencia son aspectos clave que siempre deben considerarse al elegir un enfoque de aprendizaje por refuerzo, y esta actividad me ha permitido entender mejor cómo se interrelacionan estos factores.

BIBLIOGRAFIAS

*¿Qué es el Aprendizaje mediante refuerzo? - Explicación del Aprendizaje mediante refuerzo - AWS*. (s. f.). Amazon Web Services, Inc. https://aws.amazon.com/es/what-is/reinforcement-learning/

Na, & Na. (2020, 27 diciembre). *Aprendizaje por Refuerzo | Aprende Machine Learning*. Aprende Machine Learning. https://www.aprendemachinelearning.com/aprendizaje-por-refuerzo/#:~:text=El%20aprendizaje%20por%20refuerzo%20puede,recompensando%20cuando%20lo%20hace%20bien.

Technologies, D., & Technologies, D. (2018, 2 noviembre). Inteligencia Artificial:¿Qué es el aprendizaje de refuerzo? Una Explicación Simple y Ejemplos Prácticos. Dell Technologies. https://www.dell.com/es-es/blog/inteligencia-artificial-aprendizaje-refuerzo-explicacion-y-ejemplos/

*¿Qué es el aprendizaje reforzado a partir de la retroalimentación humana (RLHF)? | IBM*. (s. f.). https://www.ibm.com/mx-es/topics/rlhf