

Centro de Ciencias Básicas

Aprendizaje Inteligente
Profesor: Dr. Francisco Javier Luna Rosas

Examen Parcial 3

Universidad Autónoma de Aguascalientes
Ingeniería en Computación Inteligente
Semestre 6° A

Integrantes:

Alegría Romero Dante Alejandro
Aranda Gonzalez Diego Alberto
Balandrán Félix Andrea Margarita
Moreno Sánchez Diego Emilio

Introducción

¿Qué es el Aprendizaje por Refuerzo?

En aprendizaje por refuerzo, se intenta hacer aprender a la máquina basándose en un esquema de “premios y castigos”, en un entorno en donde hay que tomar acciones y que está afectado por múltiples variables que cambian con el tiempo.

En Aprendizaje por Refuerzo se intenta “maximizar la recompensa”. Y esto puede ser, a pesar de a veces cometer errores o de no ser óptimos.

Componentes del Aprendizaje por Refuerzo

- El agente: será nuestro modelo que queremos entrenar y que aprende a tomar decisiones.
- Ambiente: será el entorno en donde interactúa y “se mueve” el agente. El ambiente contiene las limitaciones y reglas posibles a cada momento.

Entre ellos hay una relación que se retroalimenta y cuenta con los siguientes nexos:

- Acciones: las posibles acciones que puede tomar en un momento determinado el Agente.
- Estado (del ambiente): son los indicadores del ambiente de cómo están los diversos elementos que lo componen en ese momento.
- Recompensas (o castigos): a raíz de cada acción tomada por el Agente, podremos obtener un premio o una penalización que orientarán el Agente en si lo está haciendo bien o mal.
- Políticas (en el caso de Q – Learning): es una tabla que le indicará al modelo “como actuar” en cada estado.

En este caso utilizamos el Q – Learning, el objetivo principal al entrenar nuestro modelo a través de las simulaciones será ir “rellenando” la tabla de Políticas de manera que las decisiones que vaya tomando nuestro agente obtenga “la mayor recompensa” a la vez que avanzamos y no nos quedamos estancados, es decir, pudiendo cumplir el objetivo global (o final) que deseamos alcanzar.

Para saber cómo ir completando la tabla de políticas nos valemos de la ecuación de Bellman:

$$\hat{Q}(s,a) = Q(s,a) + \alpha \left[R + \left(\lambda \max_{a'} Q(s',a') \right) - Q(s,a) \right]$$

Diagram illustrating the components of the Q-learning equation:

- $Q(s,a)$: valor actual
- α : ratio aprendizaje
- R : recompensa
- λ : tasa descuento
- $\max Q(s',a')$: valor óptimo esperado
- $Q(s,a)$ (inside brackets): valor actual

La ecuación explica es como ir actualizando las políticas en base al valor actual más una futura recompensa que recibiremos, en caso de tomar dicha acción. Hay dos ratios que afectan a la manera en que influye esa recompensa: la ratio de aprendizaje, que regula “la velocidad” en la que se aprende y la “tasa de descuento” que tendrá en cuenta la recompensa.

Para el caso del examen tenemos el uso de Web Mining y para esta ocasión obtuvimos algunos tableros de sudokus de internet, mediante la aplicación de Q – Learning se espera que el agente resuelva uno de los tableros de la menor mas optima.

Componentes

- Agente: se encargará de resolver el tablero
 - o Sus acciones serán agregar números del 1 al 9, o cambiar el valor agregado si no se cumple con las restricciones.
- Ambiente: será el tablero de sudoku
 - o Se le establecerá limites como las restricciones que tiene el juego de sudoku.
- Estado: El estado del ambiente, puede ser el estar correcto o incorrecto la solución del tablero.
- Recompensas: se aplicará la recompensa si se cumple con las restricciones o en caso de cometer errores será un castigo, se verá afectado los valores de la q_table.

Mientras los valores de la tabla de las políticas sean positivas, los resultados podrán ser los correctos en la iteración en la que se encuentre el programa, en

todo caso si son negativas, puede que en esa iteración se encuentre errores que no generen la respuesta correcta.

Conclusiones

La implementación de estas técnicas para el Machine Learning son de gran utilidad para poder entrenar a "agentes" que puedan resolver problemas y seguir aprendiendo de estos para ser más precisos, el uso de estas técnicas es fundamentales para casos como las IA's en los videojuegos y son interesantes de aplicar.

Referencias

Na, & Na. (2020, 27 diciembre). *Aprendizaje por Refuerzo / Aprende Machine Learning*.

Aprende Machine Learning. <https://www.aprendemachinelearning.com/aprendizaje-por-refuerzo/>