Imagen que contiene Texto

Descripción generada automáticamente

**Examen\_3\_AI\_Junio\_2024**

**Aprendizaje por refuerzo Web Mining**

**Integrantes:**

Cesar Eduardo Elías del Hoyo

José Luis Sandoval Pérez

Diego Emanuel Saucedo Ortega

Carlos Daniel Torres Macías

**Universidad Autónoma de Aguascalientes**

Aguascalientes, Ags, 01 de junio, 2024

**Analisis de técnicas de ML (Aprendizaje por refuerzo) aplicadas al Web Mining.**

El Aprendizaje por Refuerzo (Reinforcement Learning, RL) es una técnica de aprendizaje automático donde un agente aprende a tomar decisiones secuenciales interactuando con un entorno. En el contexto del Web Mining, RL se puede aplicar para mejorar la experiencia del usuario, optimizar motores de búsqueda, personalizar contenidos, entre otros.

**Componentes Principales del Aprendizaje por Refuerzo**

1. **Agente:** Es el componente que toma decisiones. En Web Mining, el agente podría ser un bot de navegación web que busca optimizar ciertas métricas como el tiempo en el sitio o la tasa de conversión.
2. **Medio Ambiente (Entorno):** Es todo lo que rodea al agente y con lo que interactúa. En el contexto de Web Mining, el entorno sería la web o el sitio web específico en el que el agente está operando.
3. **Acción:** Son las decisiones que toma el agente. En Web Mining, una acción podría ser seleccionar un enlace, recomendar un producto, cambiar el diseño de una página, etc.
4. **Estado:** Representa una situación particular del entorno en un momento dado. En Web Mining, el estado podría ser una representación de la página actual, el historial de navegación del usuario, o el contenido visualizado.
5. **Algoritmo:** Es la metodología que el agente utiliza para aprender y tomar decisiones. En RL, se utilizan varios algoritmos, siendo uno de los más comunes Q-learning.
6. **Política (Policy):** Es la estrategia que el agente sigue para decidir qué acción tomar en cada estado. Puede ser determinista (una acción fija para cada estado) o estocástica (una distribución de probabilidad sobre las acciones).
7. **Función de Recompensa (Reward Function):** Es una señal de retroalimentación que indica el éxito o fracaso de una acción tomada por el agente. En Web Mining, la recompensa podría ser una métrica como el número de páginas vistas, el tiempo de permanencia en el sitio, las conversiones, etc.

En este archivo se abordará una implementación del algoritmo Q-learning, donde:

Q-learning es un algoritmo de RL off-policy que busca aprender la calidad de las acciones, que se denomina valor Q, para cada par de estado-acción. Este valor indica la recompensa esperada de tomar una acción en un estado y seguir la política óptima desde allí en adelante.

**Proceso del Q-learning**

1. **Inicialización**: Se inicializa la función Q con valores arbitrarios, generalmente ceros.
2. **Interacción con el Entorno**:
   * El agente observa el estado actual sss.
   * Elige una acción aaa basada en una política (por ejemplo, política ε-greedy que elige una acción aleatoriamente con probabilidad ε o la mejor acción conocida con probabilidad 1-ε).
   * Ejecuta la acción a y observa la recompensa r y el nuevo estado s′.
3. **Actualización de Q-valor**:
   * La fórmula de actualización es:

Imagen que contiene Gráfico

Descripción generada automáticamentedonde:

* + - α es la tasa de aprendizaje.
    - γ es el factor de descuento.
    - r es la recompensa recibida.
    - s es el nuevo estado.
    - Max a′​Q(s′,a′) es el valor Q máximo del nuevo estado s′.