Modelo de Aprendizaje por Refuerzo aplicado a un Caso de Adicción Documento Extendido

Introducción

Este documento presenta un análisis exhaustivo del caso clínico de un hombre de treinta años con consumo problemático de sustancias y dificultades sociofamiliares y laborales. El objetivo es mostrar cómo puede representarse esta situación mediante un Modelo de Decisión de Markov conocido como MDP y cómo un algoritmo de aprendizaje por refuerzo puede aprender una política que llegue racionalmente a la decisión de buscar tratamiento en una clínica privada debido a la eficacia insuficiente o lenta de los recursos públicos disponibles.

Descripción del Caso

El sujeto presenta consumo diario de marihuana y consumo de cocaína y alcohol los fines de semana. Su situación laboral es inestable. Recibe una prestación estatal y obtiene ingresos adicionales mediante trabajos ocasionales de carpintería. Convive con su pareja, quien también consume marihuana de forma ocasional, y con una hija pequeña. La pareja considera que el sujeto presenta un problema de adicción. Los intentos previos de tratamiento a través del sistema público no han dado resultados efectivos debido a listas de espera y falta de recursos asistenciales especializados.

Formulación del Problema como un MDP

Un MDP se define mediante el conjunto S de estados, el conjunto A de acciones, la función de transición T y la función de recompensa R. También incluye un factor de descuento gamma comprendido entre cero y uno. Su estructura formal es la siguiente:

MDP = (S, A, T, R, gamma)

Los estados describen la situación del sujeto en cada momento. Las acciones representan las decisiones posibles. Las transiciones modelan cómo cambia el estado tras ejecutar una acción. Las recompensas cuantifican los efectos positivos o negativos de cada acción y cada estado. El objetivo del agente es maximizar la suma esperada de recompensas descontadas.

Definición de Estados

El estado puede representarse como un vector que agrupa variables relevantes de la vida del sujeto. Cada componente puede tomar valores discretos asociados a niveles de gravedad o intensidad. Un estado puede escribirse formalmente como

$$s = (c_m, c_c, c_a, e, i, r_p, r_h, a_m, t)$$

donde

c_m representa el nivel de consumo de marihuana

c_c el nivel de consumo de cocaína

c_a el nivel de consumo de alcohol

e la situación laboral

i los ingresos aproximados

r_p la calidad de la relación de pareja

r_h la calidad de la relación con la hija

a_m el nivel de ansiedad o deterioro mental

t la situación de tratamiento

Definición de Acciones

Las acciones disponibles para el agente pueden representarse como el conjunto

$$A = \{a_1, a_2, a_3, a_4, a_5, a_6\}$$

donde

a_1 es no modificar la conducta

a_2 es intentar reducir consumo por cuenta propia

a_3 es buscar ayuda en el sistema público

a_4 es unirse a grupos de apoyo

a_5 es ingresar en una clínica privada

a_6 es abandonar tratamiento

Estas acciones cubren el abanico de decisiones realistas que una persona con esta problemática podría tomar.

Función de Transición

La función de transición T define la probabilidad de pasar del estado s al estado s prima tras ejecutar la acción a. Se expresa como

T(s, a, s') = P(s' | s, a)

En este caso las transiciones incluyen cambios en los niveles de consumo, estabilidad familiar, salud mental y progreso en tratamiento. Las transiciones no son deterministas debido a la naturaleza estocástica del comportamiento humano y de los efectos de las sustancias. Cada acción genera un conjunto de posibles estados futuros con probabilidades asignadas.

Función de Recompensa

La función de recompensa R asigna un valor numérico a cada transición o a cada estado según corresponda. Un diseño posible es

R(s, a, s') = f_beneficios(s') - f_costes(s, a)

Los beneficios incluyen mejora en salud mental y física, estabilidad familiar y funcionamiento laboral. Los costes consideran las consecuencias negativas del consumo, la pérdida de vínculos y

los gastos económicos asociados a ciertos tratamientos. Para que el modelo pueda llegar racionalmente a la decisión de utilizar una clínica privada se define que este tipo de tratamiento produce mejoras significativas y rápidas en salud y funcionamiento aunque implica un coste económico moderado.

Políticas y Valoración de Estados

Una política se define como una función

pi(s) = a

que selecciona la acción óptima para cada estado. El objetivo del aprendizaje por refuerzo es encontrar la política óptima pi_estrella que maximiza el valor esperado de retorno. El valor de un estado se define como

V_pi(s) = E [sum_{t=0}^{infinito} gamma^t R(s_t, a_t, s_{t+1}) | pi]

Para la política óptima se cumple la ecuación de Bellman

V_pi*(s) = max_a E [R(s, a, s') + gamma V_pi*(s')]

Q Learning

Una forma práctica de aprender la política óptima es mediante Q Learning. Este método aprende una función Q que aproxima el valor de tomar una acción a en un estado s. La actualización se expresa como

Q(s, a) ← Q(s, a) + alfa [R(s, a, s') + gamma max_{a'} Q(s', a') - Q(s, a)]

donde alfa es la tasa de aprendizaje. Cuando el algoritmo converge la política óptima se obtiene seleccionando la acción con mayor valor Q.

Pseudocódigo del Algoritmo

A continuación se presenta pseudocódigo claro y sin elementos gráficos innecesarios para Q Learning aplicado a este caso.

Inicializar Q(s, a) de forma arbitraria

Para cada episodio repetir

Elegir un estado inicial s

Mientras el episodio no haya terminado repetir

Seleccionar acción a siguiendo criterio epsilon codicioso

Ejecutar acción a en el entorno

Observar nuevo estado s prima y la recompensa r

Actualizar el valor Q mediante

Q(s, a) = Q(s, a) + alfa (r + gamma max Q(s prima, a prima) menos Q(s, a))

Reemplazar s por s prima

Fin

Este pseudocódigo describe el proceso mediante el cual el agente aprende por ensayo y error cómo evolucionan los estados y qué decisiones producen mayores recompensas acumuladas.

Conclusión

Al modelar adecuadamente las recompensas y transiciones se obtiene que la acción asociada al ingreso en una clínica privada produce mejoras más rápidas y sostenidas en salud física y mental y en la estabilidad familiar. Por tanto el algoritmo aprende que dicha acción maximiza la recompensa total incluso considerando su coste económico. El modelo no sustituye la intervención profesional pero constituye una herramienta formal útil para analizar escenarios complejos en el ámbito de las adicciones y de la salud mental.