Markov

La cadena de Markov, también conocida como modelo de Markov o proceso de Markov, es un concepto desarrollado dentro de la teoría de la probabilidad y la estadística que establece una fuerte dependencia entre un evento y otro suceso anterior. Su principal utilidad es el análisis del comportamiento de procesos estocásticos.

Procesos de desicion de MarKov

Los problemas de decisión las acciones adoptadas por el agente determinan no sólo la recompensa inmediata sino el siguiente estado

del entorno, al menos probabilísticamente. Por lo tanto, el agente toma en

cuenta el siguiente estado y la recompensa cuando decide tomar una acción

determinada.

Los modelos de Markov se han aplicado con éxito en la resolución de

problemas de navegación en el campo de la robótica, como ejemplos

tenemos los trabajos de Simmons & Koenig (1995), Cassandra et al. (1996),

Kaelbling et al. (1998), Koenig & Simmons (1998), Nourbakhsh et al. (1995) y

Thrun (2000).

Formalizacion

Un MDP(Markov Decision Processes) es un modelo matemático de un problema el cual explícitamente considera la incertidumbre en las acciones del sistema. La dinámica del sistema está determinada por una función de transición de probabilidad.

Políticas

MDP es un proceso secuencial de toma de decisiones, y en este sentido es posible trabajar en dos contextos distintos. En el primero de ellos, conocido como de horizonte-finito, el agente sólo actúa durante un número finito y conocido de pasos k.

Función de valor

La función de valor Vπ(s)determina la utilidad de cada estado s suponiendo que las acciones se escogen según la política π . Se trata de un concepto distinto al de recompensa. La función de recompensa asigna, para cada una de las acciones que pueden ejecutarse en cada estado, un valor numérico 206 que representa la utilidad inmediata de dicha acción.

Políticas optimas

Resolver un MDP consiste en encontrar la política óptima, una directiva de control que maximiza la función de valor sobre los estados. Teóricamente, es posible obtener todas las posibles políticas para un MDP y a continuación escoger entre ellas, aquella que maximiza la función de valor.

Proceso de decisión de Markov en tiempo continuo

En los procesos de decisión de Markov de tiempo discreto, las decisiones se toman en intervalos de tiempo discretos. Sin embargo, para los procesos de decisión de Markov de tiempo continuo , las decisiones se pueden tomar en cualquier momento que elija el tomador de decisiones.

Si el espacio de estados y el espacio de acción son finitos.

$\mathcal{S}$: Espacio de Estados;
$\mathcal{A}$: Espacio de acción;
$q(i \mid j, a)$: , función de tasa de transición;$\mathcal{S} \times \mathcal{A} \to \triangle \mathcal{S}$
$R(i, a)$: , una función de recompensa.$\mathcal{S} \times \mathcal{A} \to \mathbb{R}$

Si el espacio de estado y el espacio de acción son continuos.

$\mathcal{X}$: espacio de Estados;
$\mathcal{U}$: espacio de posible control;
$f(x, u)$: , una función de tasa de transición;$\mathcal{X} \times \mathcal{U} \to \triangle \mathcal{X}$
$r(x, u)$: , una función de tasa de recompensa tal que , donde está la función de recompensa que discutimos en el caso anterior.$\mathcal{X} \times \mathcal{U} \to \mathbb{R} r(x(t), u(t))\, dt = dR(x(t), u(t)) R(x, u)$

Problema

Al igual que los procesos de decisión de Markov de tiempo discreto, en los procesos de decisión de Markov de tiempo continuo queremos encontrar la política o el control óptimos que podrían darnos la recompensa integrada óptima esperada.

$$\max \mathrm{E}_u \left[\int_0^\infty \gamma^t r(x(t), u(t))\, dt \,\middle|\, x_0 \right]$$

dónde $0 \leq \gamma < 1$.

Formulación de programación lineal.

Programa lineal primario (P-LP).

$$\begin{aligned} \text{Minimize} \quad & g \\ \text{s.t} \quad & g - \sum_{j \in S} q(j \mid i,a) h(j) \geq R(i,a) \ \forall i \in S,\ a \in A(i) \end{aligned}$$

Programa lineal dual (D-LP).

$$\begin{aligned} \text{Maximize} & \sum_{i \in S} \sum_{a \in A(i)} R(i,a) y(i,a) \\ \text{s.t.} & \sum_{i \in S} \sum_{a \in A(i)} q(j \mid i,a) y(i,a) = 0 \quad \forall j \in S, \\ & \sum_{i \in S} \sum_{a \in A(i)} y(i,a) = 1, \\ & y(i,a) \geq 0 \qquad \forall a \in A(i) \text{ and } \forall i \in S \end{aligned}$$

Tecnicas que existen para resolver estos procesos de Markov

Ejemplo de Retorno de la Cadena de Markov(Funcion Valor)

- Estados: {calor, frío, lluvia} (supondremos que

s1 = calor, s2 = frío y s3 = lluvia)

- Matriz de probabilidades de transición:

a11 = P(Xt = calor|Xt−1 = calor) = 0.5

a12 = P(Xt = frío|Xt−1 = calor) = 0.2

a13 = P(Xt = lluvia|Xt−1 = calor) = 0.3

a21 = P(Xt = calor|Xt−1 = frío) = 0.2

a22 = P(Xt = frío|Xt−1 = frío) = 0.5

a23 = P(Xt = lluvia|Xt−1 = frío) = 0.3

a31 = P(Xt = calor|Xt−1 = lluvia) = 0.3

a32 = P(Xt = frío|Xt−1 = lluvia) = 0.1

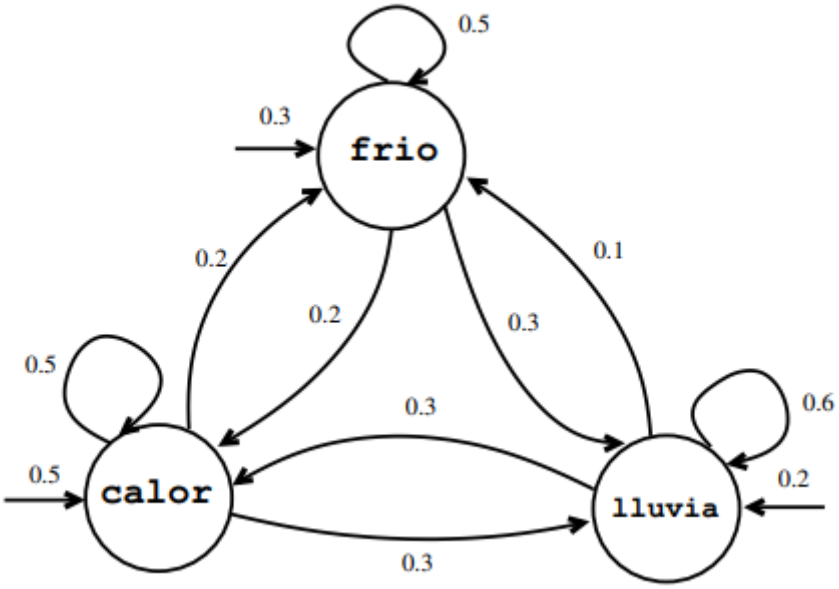
a33 = P(Xt = lluvia|Xt−1 = lluvia) = 0.6

- Probabilidades iniciales:

π1 = P(X1 = calor) = 0.5

π2 = P(X1 = frío) = 0.3

π3 = P(X1 = lluvia) = 0.2



- Los nodos son los estados
- Los arcos están etiquetados con la probabilidad de pasar de un estado a otro

Ejemplo de Formalizacion

La metodología adoptada permite enfrentar la incertidumbre presente en esta clase de problemas, describiendo la dinámica de la permanencia de los pacientes en términos probabilísticos.

El modelo empleado resulta satisfactorio para abordar el problema en estudio, tomando en cuenta su comportamiento y las diferencias obtenidas en comparación con la muestra de datos con la que se contrastó.

Un aumento en la cantidad de estados, para diferenciar de manera menos agregada los niveles de gravedad, como la definición de una etapa por periodos diarios o de medios días, pueden ser

fácilmente incorporados sin alterar mayormente el modelo adoptado ni la complejidad del mismo.

A continuacion se presentan, en forma breve, los elementos que integran un proceso markoviano de decision, abreviado mediante.

Para simplificar la exposici´on se supone que el espacio de estados, S , es discreto. Considere una cadena de Markov controlada en tiempo

discreto con:

i) Un espacio de estados, finito o numerable S.

ii) Un espacio medible de acciones, A, equipado con una σ-´algebra

A de A. En este caso, el conjunto de restricciones se representa

mediante $K = S \times A$.

iii) Para cada estado $i \in S$ existe un conjunto de acciones $A(i)$ disponibles. Estos conjuntos se suponen elementos de A.

iv) Una matriz de probabilidades de transici´on $[q(j \mid i, a)]$. Para cada

$i, j \in S$ y $a \in A(i)$ la funcion $q(j \mid i, a)$ es no negativa y medible,

$q(j \mid i, a) = 1$ para cada $i \in S$ y $a \in A(i)$.

v) Una funcion $r : K \rightarrow R$, llamada la utilidad, ganancia o costo,

dependiendo del contexto.