upper_confidence

Daniel Tello

30-04-2020

Upper confidence

El problema del bandido multibrazo

El problema es descubrir cual es la distribución de cada máquina tragamonedas. Es necesario conbinar la aplicación de estas máquinas.

- Tenemos **d** brazos. Por ejemplo los brazos son anuncios que mostramos a los usarios, cuando se conectan a una página web.

Algoritmo Upper confidence Bound

- Cada vez que un usuario se conecta a la página web, se desencadena una ronda.
- En cada ronda, n, se elige uno de los anuncios a ser mostrados por el usuario.
- A cada ronda n, el anuncio i da una recompensa: $r_i(n) \in \{0, 1\}$, si $r_i(n) = 1$, el usuario hace click en el anuncio i en la ronda n, $r_i(n) = 0$ $, en caso contrario.
- Objetivo: maximizar la recompensa a través de las rondas que se llevan a cabo.

Paso 1

A cada ronda n, se consideran dos números para cada i.

- $N_i(n)$: El número de veces que el anuncio i se selecciona en la ronda n.
- $R_i(n)$: La suma de las recompensas del anuncio i hasta la ronda n.

Paso 2

A partir de estos dos números calculamos.

- La recompensa media del anuncio i hasta la ronda n.

$$\bar{r_i}(n) = \frac{R_i(n)}{N_i(n)}$$

- El intervalo de confianza en la ronda **n**: $(\bar{r_i} - \Delta_i(n), \bar{r_i}(n) + \Delta_i(n))$ con $\Delta_i(n) = \sqrt{\frac{3\log(n)}{2N_i(n)}}$

Paso 3

Se selecciona el anuncio i con mayor límite superior del intervalo de confianza (UCB).