

Juegos iterativos en grafos para el cuidado de reservas ecológicas.

Nicolás Betancourt
Mauricio Velasco
Universidad de los Andes (Colombia)

Escuela CICADA
Octubre 2022
Punta del Este, Uruguay

Plan de la charla:

- ① Motivación: La protección de Jama Coaque.
- ② Aprendizaje sobre la marcha (multi-armed bandits)
- ③ Un algoritmo para el problema del guardabosques
(combinatorial multi-armed bandits).
- ④ Resultados teóricos y ejemplos computacionales.
- ⑤ Bandidos adversarios y desarrollos futuros.

Motivación:

Mitigar el impacto negativo de la actual pérdida de biodiversidad es uno de los retos más significativos de los años venideros.

Motivación:

Mitigar el impacto negativo de la actual pérdida de biodiversidad es uno de los retos más significativos de los años venideros.

Un mecanismo esencial de mitigación es el **mantenimiento de áreas protegidas de gran tamaño** pues estas sirven como hábitats para una amplia variedad de especies así como de reservorios de agua.

Motivación:

Desafortunadamente las áreas protegidas alrededor del mundo están constantemente bajo amenazas: **caza, tala y minería ilegales y tráfico de especies entre otras.**

Motivación:

Desafortunadamente las áreas protegidas alrededor del mundo están constantemente bajo amenazas: **caza, tala y minería ilegales y tráfico de especies entre otras.**

Esto es un reto para aquellos que cuidan estas áreas pues deben **asignar los limitados recursos a su disposición para el cuidado de zonas extensas.** Los cuidadores están típicamente en desventaja considerable respecto a los atacantes.

Ejemplo:

La reserva de **Jama Coaque** en la costa del Ecuador es un área protegida de 600 hectáreas ($3\% \times$ MVD) de *bosque húmedo tropical*. La reserva se creó en 2007 y ha ido expandiéndose de su extensión inicial de 95 hectáreas (Third Millenium alliance TMA).

Ejemplo:

La reserva de **Jama Coaque** en la costa del Ecuador es un área protegida de 600 hectáreas ($3\% \times$ MVD) de *bosque húmedo tropical*. La reserva se creó en 2007 y ha ido expandiéndose de su extensión inicial de 95 hectáreas (Third Millennium alliance TMA).

De wikipedia: "The Jama-Coaque Ecological Reserve serves as habitat and key migratory channel for six endangered species of felines (jaguar, puma, ocelot, oncilla, margay, and jaguarundi) and two endangered species of primates (mantled howler monkey and white-fronted capuchin monkey). Other endangered mammals include the tayra, the three-toed sloth, the western agouti, and the spotted paca. In 2009, herpetologist Paul S. Hamilton discovered two new species of frog in the cloud forest of the Jama-Coaque Ecological Reserve".

Jama Coaque:



Jama Coaque:



Jama Coaque:

El terreno es tan difícil que **tanto los guardabosques como los cazadores** se desplazan solamente a lo largo de un **grafo de caminos** establecido.

Jama Coaque:

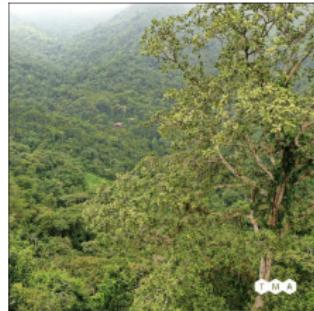
El terreno es tan difícil que **tanto los guardabosques como los cazadores** se desplazan solamente a lo largo de un **grafo de caminos** establecido.

En la reserva hay 138 **senderos** (aristas del grafo). Los guardabosques empiezan en la casa de Bambú y recorren la reserva caminando en ciclos.

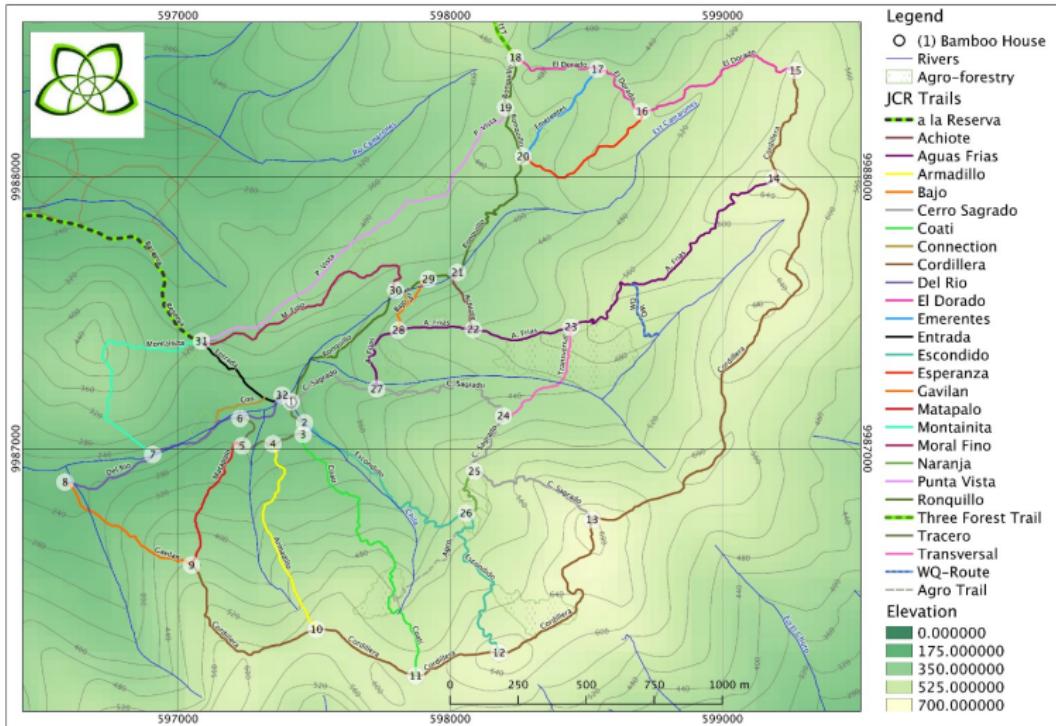
Jama Coaque:

El terreno es tan difícil que **tanto los guardabosques como los cazadores** se desplazan solamente a lo largo de un **grafo de caminos** establecido.

En la reserva hay 138 **senderos** (aristas del grafo). Los guardabosques empiezan en la casa de Bambú y recorren la reserva caminando en ciclos. La reserva tiene 135.010 **ciclos simples** (de los cuales alrededor de 10.000 pasan por la casa de Bambú).



Jama Coaque:



Jama Coaque:

Los guardabosques nos dicen que la reserva esta afectada principalmente por **tala ilegal** de balsos y **caza** de zainos y venados.

Jama Coaque:

Los guardabosques nos dicen que la reserva esta afectada principalmente por **tala ilegal** de balsos y **caza** de zainos y venados. La reserva esta protegida por un grupo de 4 guardabosques que patrullan la reserva diariamente a lo largo de ciclos distintos.

Problema:

En esta charla nos ocuparemos de los siguientes dos problemas:

- Cómo diseñar **agendas de patrullado** que permitan que los guardabosques descubran (y frustren) la mayor cantidad de crímenes ambientales posible?
- Cómo almacenar y utilizar la información disponible en la reserva para que los guardabosques puedan tomar mejores decisiones **en tiempo real?**

Problema:

En esta charla nos ocuparemos de los siguientes dos problemas:

- Cómo diseñar **agendas de patrullado** que permitan que los guardabosques descubran (y frustren) la mayor cantidad de crímenes ambientales posible?
- Cómo almacenar y utilizar la información disponible en la reserva para que los guardabosques puedan tomar mejores decisiones **en tiempo real?**

El problema es difícil porque: hay 10.000 ciclos válidos en la reserva, hay sólo 4 guardabosques y la localización, cantidad y rutas de los cazadores es desconocida.

Problema:

En esta charla nos ocuparemos de los siguientes dos problemas:

- Cómo diseñar **agendas de patrullado** que permitan que los guardabosques descubran (y frustren) la mayor cantidad de crímenes ambientales posible?
- Cómo almacenar y utilizar la información disponible en la reserva para que los guardabosques puedan tomar mejores decisiones **en tiempo real?**

El problema es difícil porque: hay 10.000 ciclos válidos en la reserva, hay sólo 4 guardabosques y la localización, cantidad y rutas de los cazadores es desconocida.

*Es necesario combinar la exploración constante de la reserva con visitar los lugares donde sabemos que es más probable que haya crímenes ambientales (**explotación** del conocimiento adquirido).*

APRENDIZAJE SOBRE LA MARCHA (online learning).

Multi-armed bandits:

*Un jugador tiene un conjunto de m máquinas tragamonedas numeradas $1, 2, \dots, m$ (**brazos**). En cada turno el jugador escoge una máquina, hala el brazo correspondiente y obtiene una cierta cantidad de dinero. **Qué estrategia debe seguir el jugador para maximizar su retorno en T turnos?***



Multi-armed bandits:

Si supiéramos las medias μ_i este problema sería muy fácil.

Multi-armed bandits:

Si supiéramos las medias μ_i este problema sería muy fácil.

Buscamos la mejor máquina $j^ := \operatorname{argmax}_{j \in [m]} (\mu_j)$ y usamos esa todos los turnos.*

Multi-armed bandits:

Si supiéramos las medias μ_i este problema sería muy fácil.

Buscamos la mejor máquina $j^ := \operatorname{argmax}_{j \in [m]} (\mu_j)$ y usamos esa todos los turnos.*

El problema es que **el jugador desconoce las medias**. Así que debe usar algo de su tiempo para intentar aprender cuáles máquinas tienen buen retorno y otra parte de su tiempo para jugar en estas máquinas.

Multi-armed bandits:

Si supiéramos las medias μ_i este problema sería muy fácil.

Buscamos la mejor máquina $j^ := \operatorname{argmax}_{j \in [m]} (\mu_j)$ y usamos esa todos los turnos.*

El problema es que **el jugador desconoce las medias**. Así que debe usar algo de su tiempo para intentar aprender cuáles máquinas tienen buen retorno y otra parte de su tiempo para jugar en estas máquinas.

El multi-armed bandit es un problema fundamental pues abstrae el dilema entre exploración y explotación.

Multi-armed bandits:

Un **multi-armed bandit problem** con m brazos esta determinado por la sucesión doble de variables aleatorias (retornos) $X_{i,n}$ donde i es el índice del brazo y n el turno en el que nos encontramos.

Multi-armed bandits:

Un **multi-armed bandit problem** con m brazos esta determinado por la sucesión doble de variables aleatorias (retornos) $X_{i,n}$ donde i es el índice del brazo y n el turno en el que nos encontramos.

Asumimos:

- Independencia de un brazo con su pasado y con los demás brazos.
- Que el soporte de los retornos esta contenido en el intervalo $[0, 1]$.
- Que la distribución de $X_{i,n}$ es la misma para todo n .

y querríamos escoger brazos sucesivamente para maximizar el retorno esperado (basandonos sólamente en los retornos observados).

Estimación de medias:

Si halamos n veces un mismo brazo, los retornos de ese brazo nos dan una sucesión i.i.d. Z_1, \dots, Z_n con media μ y soporte en un intervalo.

Estimación de medias:

Si halamos n veces un mismo brazo, los retornos de ese brazo nos dan una sucesión i.i.d. Z_1, \dots, Z_n con media μ y soporte en un intervalo.

Si definimos la media muestral

$$S_n := \frac{1}{n} (Z_1 + \cdots + Z_n)$$

entonces $\mathbb{E}[S_n] = \mu$ y además

Lema. (Ley débil de grandes números)

Para todo $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P} \{|S_n - \mu| > \epsilon\} = 0$$

Estimación de medias:

Si halamos n veces un mismo brazo, los retornos de ese brazo nos dan una sucesión i.i.d. Z_1, \dots, Z_n con media μ y soporte en un intervalo.

Si definimos la media muestral

$$S_n := \frac{1}{n} (Z_1 + \cdots + Z_n)$$

entonces $\mathbb{E}[S_n] = \mu$ y además

Lema. (Ley débil de grandes números)

Para todo $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P} \{|S_n - \mu| > \epsilon\} = 0$$

Es decir $S_n - \epsilon \leq \mu \leq S_n + \epsilon$ con **gran probabilidad** para n lo suficientemente grande.

Estimación de medias:

Teorema. (Desigualdad de Hoeffding)

Sean Z_1, \dots, Z_n variables aleatorias independientes tomando valores en $[0, 1]$ y con media μ . Entonces

$$\mathbb{P}\{|S_n - \mu| \geq t\} \leq 2 \exp(-2nt^2)$$

Estimación de medias:

Teorema. (Desigualdad de Hoeffding)

Sean Z_1, \dots, Z_n variables aleatorias independientes tomando valores en $[0, 1]$ y con media μ . Entonces

$$\mathbb{P}\{|S_n - \mu| \geq t\} \leq 2 \exp(-2nt^2)$$

Es decir, dado $\epsilon > 0$ y una cota β para la probabilidad podemos encontrar $n(\epsilon, \beta)$ tal que

$$S_n - \epsilon \leq \mu \leq S_n + \epsilon$$

con **probabilidad** por lo menos β para todo $n \geq n(\epsilon, \beta)$.

Soluciones de calidad

Una **estrategia** A es un algoritmo que nos dice qué brazo escoger en cada turno basado en los retornos de jugadas anteriores.

Soluciones de calidad

Una **estrategia** A es un algoritmo que nos dice qué brazo escoger en cada turno basado en los retornos de jugadas anteriores.

Sea $T_i(n)$ el número de veces que la máquina i ha sido seleccionada durante los primeros n turnos.

Soluciones de calidad

Una **estrategia** A es un algoritmo que nos dice qué brazo escoger en cada turno basado en los retornos de jugadas anteriores.

Sea $T_i(n)$ el número de veces que la máquina i ha sido seleccionada durante los primeros n turnos.

Definición.

*El arrepentimiento **regret** de A después de n jugadas se define como la pérdida de ganancias promedio que resulta de usar la estrategia A y no la estrategia óptima.*

es decir:

$$R(n) := \mu^* n - \sum_{j=1}^m \mu_j \mathbb{E}[T_j(n)]$$

dónde $\mu^* := \max_{i \in [m]} \mu_i$.

Soluciones de calidad

Una **estrategia** A es un algoritmo que nos dice qué brazo escoger en cada turno basado en los retornos de jugadas anteriores.

Sea $T_i(n)$ el número de veces que la máquina i ha sido seleccionada durante los primeros n turnos.

Definición.

El arrepentimiento regret de A después de n jugadas se define como la pérdida de ganancias promedio que resulta de usar la estrategia A y no la estrategia óptima.

es decir:

$$R(n) := \mu^* n - \sum_{j=1}^m \mu_j \mathbb{E}[T_j(n)]$$

dónde $\mu^* := \max_{i \in [m]} \mu_i$.

El regret siempre es positivo y **nuestro objetivo es hacerlo lo más pequeño posible**.

Upper confidence bound policy (UCB):

En 2002 Auer, Cesa-Bianchi y Fischer proponen la siguiente política:

- *Inicialización: Juegue cada brazo una vez.*
- *Después, en cada turno $n \geq m + 1$:*
 - ① *Calculamos el índice j que maximiza la cantidad*

$$\bar{x}_j + \sqrt{\frac{2 \ln(n)}{n_j}}$$

dónde \bar{x}_j es el retorno promedio obtenido por el j -ésimo brazo hasta ahora y n_j es el número de veces que el j -ésimo brazo ha sido usado hasta ahora.

- ② *Tiramos del brazo j . Anotamos los retornos y actualizamos nuestras estimaciones.*

Teorema. (Auer, Cesa-Bianchi, Fischer, 2002)

Para todo $m > 1$ y distribuciones de retorno con soporte en $[0, 1]$, la política UCB satisface la desigualdad

$$R(n) \leq \left[8 \sum_{i:\mu_i < \mu^*} \frac{\log(n)}{\Delta_i} \right] + \left(1 + \frac{\pi^2}{3} \right) \left(\sum_{i=1}^m \Delta_i \right)$$

donde $\Delta_i := \mu^* - \mu_i$

Teorema. (Auer, Cesa-Bianchi, Fischer, 2002)

Para todo $m > 1$ y distribuciones de retorno con soporte en $[0, 1]$, la política UCB satisface la desigualdad

$$R(n) \leq \left[8 \sum_{i:\mu_i < \mu^*} \frac{\log(n)}{\Delta_i} \right] + \left(1 + \frac{\pi^2}{3} \right) \left(\sum_{i=1}^m \Delta_i \right)$$

donde $\Delta_i := \mu^* - \mu_i$

En particular, tenemos que

$$\lim_{N \rightarrow \infty} \frac{R(N)}{N} = 0$$

Teorema. (Auer, Cesa-Bianchi, Fischer, 2002)

Para todo $m > 1$ y distribuciones de retorno con soporte en $[0, 1]$, la política UCB satisface la desigualdad

$$R(n) \leq \left[8 \sum_{i:\mu_i < \mu^*} \frac{\log(n)}{\Delta_i} \right] + \left(1 + \frac{\pi^2}{3} \right) \left(\sum_{i=1}^m \Delta_i \right)$$

donde $\Delta_i := \mu^* - \mu_i$

La demostración del teorema consiste en verificar que para toda máquina subóptima se tiene $\mathbb{E}[T_j(n)] \leq \frac{8 \log(n)}{\Delta_j^2}$ mediante la desigualdad de Hoeffding.

Teorema. (Auer, Cesa-Bianchi, Fischer, 2002)

Para todo $m > 1$ y distribuciones de retorno con soporte en $[0, 1]$, la política UCB satisface la desigualdad

$$R(n) \leq \left[8 \sum_{i: \mu_i < \mu^*} \frac{\log(n)}{\Delta_i} \right] + \left(1 + \frac{\pi^2}{3} \right) \left(\sum_{i=1}^m \Delta_i \right)$$

donde $\Delta_i := \mu^* - \mu_i$

La demostración del teorema consiste en verificar que para toda máquina subóptima se tiene $\mathbb{E}[T_j(n)] \leq \frac{8 \log(n)}{\Delta_j^2}$ mediante la desigualdad de Hoeffding.

Se sabe (Lai y Robbins, 1985) que **para cualquier estrategia** se tiene la desigualdad

$$\mathbb{E}[T_j(n)] \geq \frac{\log(n)}{D(p_j \| p^*)}$$

así que el resultado de arriba es asintóticamente óptimo.

EL PROBLEMA DEL GUARDABOSQUES

De regreso a Jama Coaque:

Podemos pensar el problema del guardabosques como un MAB?

De regreso a Jama Coaque:

Podemos pensar el problema del guardabosques como un MAB?
En primera instancia si. Cada día el guardabosques debe escoger un ciclo con la intención de encontrar en cuáles de estos hay actividad ilegal.

De regreso a Jama Coaque:

Podemos pensar el problema del guardabosques como un MAB? En primera instancia si. Cada día el guardabosques debe escoger un ciclo con la intención de encontrar en cuáles de estos hay actividad ilegal.

No obstante, este acercamiento tiene problemas:

- Hay demasiadas medias que estimar (si asignamos una a cada ciclo).
- Al pensar los ciclos como cajas negras independientes hay mucha información que estamos desperdiciando (por ejemplo que diferentes ciclos comparten aristas).
- Pensar que los ciclos son entre sí independientes es una suposición muy extraña para ciclos muy parecidos.
- Querríamos poder asignar más guardabosques en cada turno.

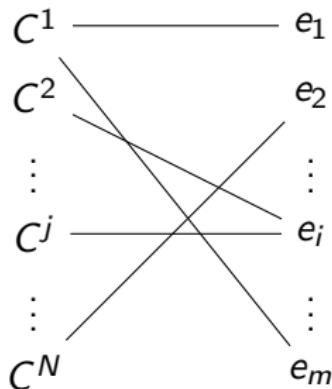
De regreso a Jama Coaque:

Proponemos entonces el siguiente modelo:

- ① La actividad ilegal es una función de las condiciones del terreno (localización de especies animales y vegetales, cercanía a fuentes de agua, etc.). Postulamos que tales actividades ocurren en cada arista según una Bernoulli con parámetro $p(e)$ (que se samplean independientemente cada día y son independientes para diferentes aristas).
- ② Cada día, un conjunto de B guardabosques elige qué ciclos patrullar. Los guardabosques quieren descubrir actividad ilegal en la reserva pero también visitar aquellos lugares donde saben por experiencia que la actividad ilegal es más probable.

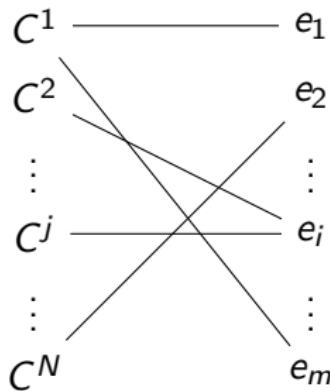
El jugador (el equipo de guardabosques) activa **un conjunto de brazos del problema en cada turno** (todas las aristas de los B ciclos seleccionados) y anota los datos observados (en cuáles de las aristas visitadas hay actividad ilegal y en cuáles no) para su uso posterior.

De regreso a Jama Coaque:



Cada día el jugador (el equipo de guardabosques) selecciona un conjunto S de exactamente B ciclos, los explora (activando un conjunto de brazos del problema en cada turno) y anota los datos observados. El retorno $r(S)$ de la escogencia S es igual al número de aristas cubiertas por S con actividad ilegal.

De regreso a Jama Coaque:



Cada día el jugador (el equipo de guardabosques) selecciona un conjunto S de exactamente B ciclos, los explora (activando un conjunto de brazos del problema en cada turno) y anota los datos observados. El retorno $r(S)$ de la escogencia S es igual al número de aristas cubiertas por S con actividad ilegal.

$$r(S) = \sum_{e \in \bigcup S} \chi_{\text{ActividadIllegal}}(e)$$

Combinatorial multi-armed bandits:

Este tipo de problemas se llaman **Combinatorial multi-armed bandits (CMAB)** pues en cada turno se elige un super-brazo (definido como un subconjunto de los m brazos disponibles) y la selección del superbrazo típicamente involucra un problema de **optimización combinatoria**.

Combinatorial multi-armed bandits:

Este tipo de problemas se llaman **Combinatorial multi-armed bandits (CMAB)** pues en cada turno se elige un super-brazo (definido como un subconjunto de los m brazos disponibles) y la selección del superbrazo típicamente involucra un problema de **optimización combinatoria**.

Hay una teoría de CMAB análoga a la de MAB. La diferencia principal radica en que los problemas de optimización combinatoria que hay que resolver típicamente no pueden resolverse de manera exacta sino sólamente aproximada (en tiempo polinomial). Esto lleva a redefinir el concepto de **regret como arrepentimiento comparado con el óptimo alcanzable algorítmicamente con toda la información probabilística disponible** (y no comparado con el óptimo ideal) y permite un desarrollo teórico muy conveniente.

Conociendo las probabilidades: Problemas de cubrimiento

Si conociéramos las probabilidades $p(e)$ entonces el retorno esperado de escoger un conjunto de ciclos S es

$$r_p(S) := \mathbb{E}[r(S)] = \sum_{e \in \bigcup S} p(e)$$

Conociendo las probabilidades: Problemas de cubrimiento

Si conociéramos las probabilidades $p(e)$ entonces el retorno esperado de escoger un conjunto de ciclos S es

$$r_p(S) := \mathbb{E}[r(S)] = \sum_{e \in \bigcup S} p(e)$$

La escogencia de ciclos se vuelve una variante del **problema de cubrimiento con pesos (weighted coverage problem)**, que nos pide escoger una colección S de B conjuntos a la izquierda cuya unión tenga actividad ilegal esperada máxima.

Conociendo las probabilidades: Problemas de cubrimiento

Si conociéramos las probabilidades $p(e)$ entonces el retorno esperado de escoger un conjunto de ciclos S es

$$r_p(S) := \mathbb{E}[r(S)] = \sum_{e \in \bigcup S} p(e)$$

La escogencia de ciclos se vuelve una variante del **problema de cubrimiento con pesos (weighted coverage problem)**, que nos pide escoger una colección S de B conjuntos a la izquierda cuya unión tenga actividad ilegal esperada máxima.

- Este problema es **NP-hard** así que incluso conociendo las probabilidades el problema del guardabosques no es fácil.
- Adicionalmente nuestro problema es ligeramente diferente pues **queremos contar el peso de un sendero varias veces si este sendero esta cubierto por varios ciclos**, porque nuestro objetivo es maximizar el número de encuentros entre guardabosques y cazadores.

Una formulación entera con relajación aleatoria

Suponga que hay N ciclos permitidos para los guardabosques y un total de m senderos y sea \mathcal{C} la matriz de incidencia aristas-ciclos (de $m \times N$). Queremos encontrar $\eta \in \{0, 1\}^N$ de tal manera que el número esperado de encuentros sea máximo.

Una formulación entera con relajación aleatoria

Suponga que hay N ciclos permitidos para los guardabosques y un total de m senderos y sea \mathcal{C} la matriz de incidencia aristas-ciclos (de $m \times N$). Queremos encontrar $\eta \in \{0, 1\}^N$ de tal manera que el número esperado de encuentros sea máximo.

Para ello resolvemos el siguiente problema de programación entera

$$\max_{(\eta, y) \in \mathbb{R}^N \times \mathbb{R}^m} \sum_{e \in E} y_e p(e) \text{ sujeto a}$$

$$\begin{aligned} \mathcal{C}\eta = y, \sum_{j=1}^N \eta_j &= B \\ \eta_j &\in \{0, 1\}, y_e \in \mathbb{N}. \end{aligned}$$

Una formulación entera con relajación aleatoria

$$\max_{(\eta, y) \in \mathbb{R}^N \times \mathbb{R}^m} \sum_{e \in E} y_e p(e) \text{ sujeto a}$$

$$\begin{aligned} \mathcal{C}\eta = y, \quad & \sum_{j=1}^N \eta_j = B \\ \eta_j \in \{0, 1\}, \quad & y_e \in \mathbb{N}. \end{aligned}$$

Una formulación entera con relajación aleatoria

$$\max_{(\eta, y) \in \mathbb{R}^N \times \mathbb{R}^m} \sum_{e \in E} y_e p(e) \text{ sujeto a}$$

$$\begin{aligned} \mathcal{C}\eta = y, \quad & \sum_{j=1}^N \eta_j = B \\ \eta_j \in \{0, 1\}, \quad & y_e \in \mathbb{N}. \end{aligned}$$

Este problema tiene una relajación continua natural, reemplazando la última línea por $\eta_j \in [0, 1]$, $y_e \in [0, B]$.

Resolviendo ese problema lineal y normalizando el óptimo obtenemos una distribución de probabilidad η_j^*/B que al samplearla sin reemplazo B -veces nos da una colección de ciclos.

Una formulación entera con relajación aleatoria

$$\max_{(\eta, y) \in \mathbb{R}^N \times \mathbb{R}^m} \sum_{e \in E} y_e p(e) \text{ sujeto a}$$

$$\begin{aligned} \mathcal{C}\eta = y, \quad & \sum_{j=1}^N \eta_j = B \\ \eta_j \in \{0, 1\}, \quad & y_e \in \mathbb{N}. \end{aligned}$$

Este problema tiene una relajación continua natural, reemplazando la última línea por $\eta_j \in [0, 1]$, $y_e \in [0, B]$.

Resolviendo ese problema lineal y normalizando el óptimo obtenemos una distribución de probabilidad η_j^*/B que al samplearla sin reemplazo B -veces nos da una colección de ciclos.

Teorema. (Betancourt, -)

Esta colección de ciclos tiene un peso de por lo menos $(1 - 1/e)OPT$ donde OPT es el óptimo del problema entero.

Aprendiendo las probabilidades:

El problema ahora es que **en la práctica los guardabosques no conocen las probabilidades**. Para aprenderlas usaremos CUCB:

- *Inicialización: Recorra ciclos que cubran la reserva una vez.*
- *Después, en cada turno n posterior:*
 - 1 *Calculamos nuestras estimaciones p_e de la probabilidad de actividad ilegal en cada arista*

$$\hat{p}_e(n) := \min \left(\overline{p_e} + \sqrt{\frac{3 \ln(n)}{2n_e}}, 1 \right)$$

dónde $\overline{p_e}$ es el promedio de actividad ilegal vista en la arista e hasta ahora y n_e es el número de veces que hemos visitado la arista e hasta ahora.

- 2 *Los guardabosques visitan los ciclos determinados por nuestro algoritmo de aproximación asumiendo que los pesos son los $(\hat{p}_e(n))_e$ estimados en (1).*

Regret aproximado:

Definición.

*Definimos el **regret** de una estrategia cualquiera A como*

$$R(n) = n(1 - 1/e)OPT(p) - \sum_{t=1}^n \mathbb{E}_p [r(S_t^A)]$$

Regret aproximado:

Definición.

Definimos el **regret** de una estrategia cualquiera A como

$$R(n) = n(1 - 1/e)OPT(p) - \sum_{t=1}^n \mathbb{E}_p [r(S_t^A)]$$

Observaciones:

- El regret depende del vector (desconocido) de probabilidades reales p .
- Comparamos el óptimo que podríamos garantizar prácticamente conociendo las probabilidades (mediante un algoritmo aproximado eficiente) con lo que nuestro algoritmo produce (en promedio, sin conocer las probabilidades).

Una garantía de éxito:

Definimos el conjunto de superbrazos malos

$$\mathcal{B} := \{S : r_p(S) < (1 - 1/e) Opt(p)\}$$

y usándolo definimos los parámetros

$$\Delta_{\min} := \min_{S \in \mathcal{B}} [(1 - 1/e) Opt(p) - r_p(S)]$$

$$\Delta_{\max} := \max_{S \in \mathcal{B}, e \in S} [(1 - 1/e) Opt(p) - r_p(S)]$$

Usando las ideas de Chen, Wang y Yuan (2016) demostramos el siguiente

Teorema. (Betancourt, -)

El algoritmo CUCB propuesto satisface

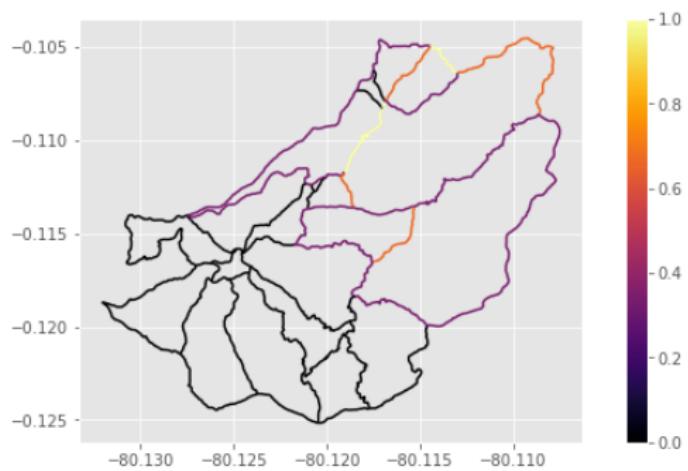
$$R(n) \leq \left[\frac{6 \log(n)}{\Delta_{\min}^2} + \frac{\pi^2}{3} + 1 \right] m \Delta_{\max}$$

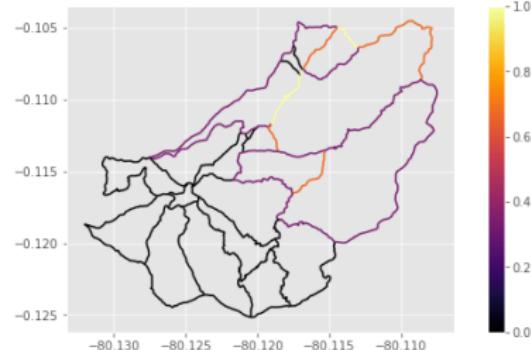
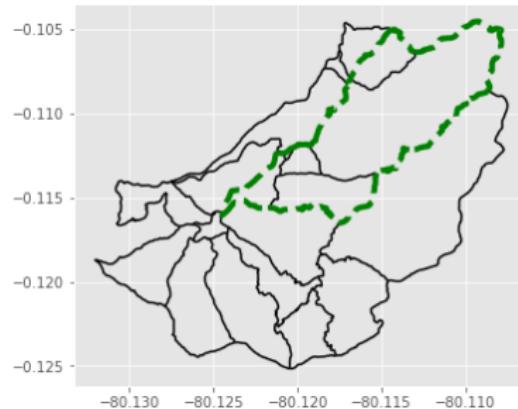
donde m es el número de senderos.



Ejemplo:

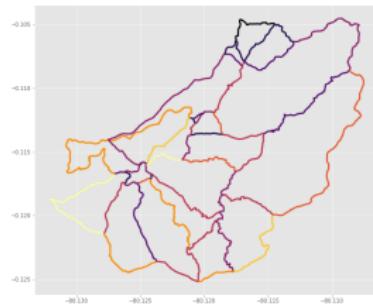
Rutas del cazador



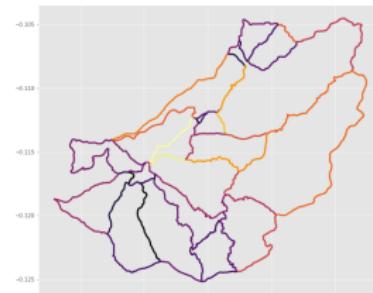


Frecuencias de patrullado de CUCB

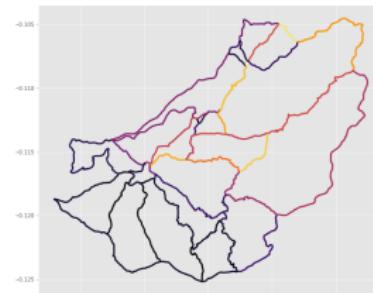
$t = 10$



$t = 50$

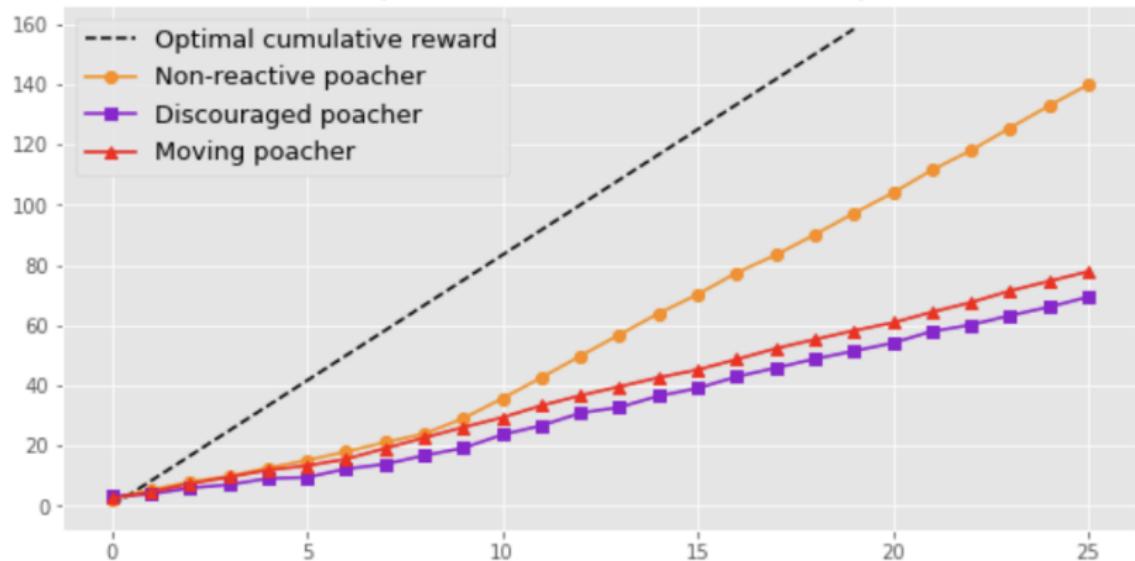


$t = 180$



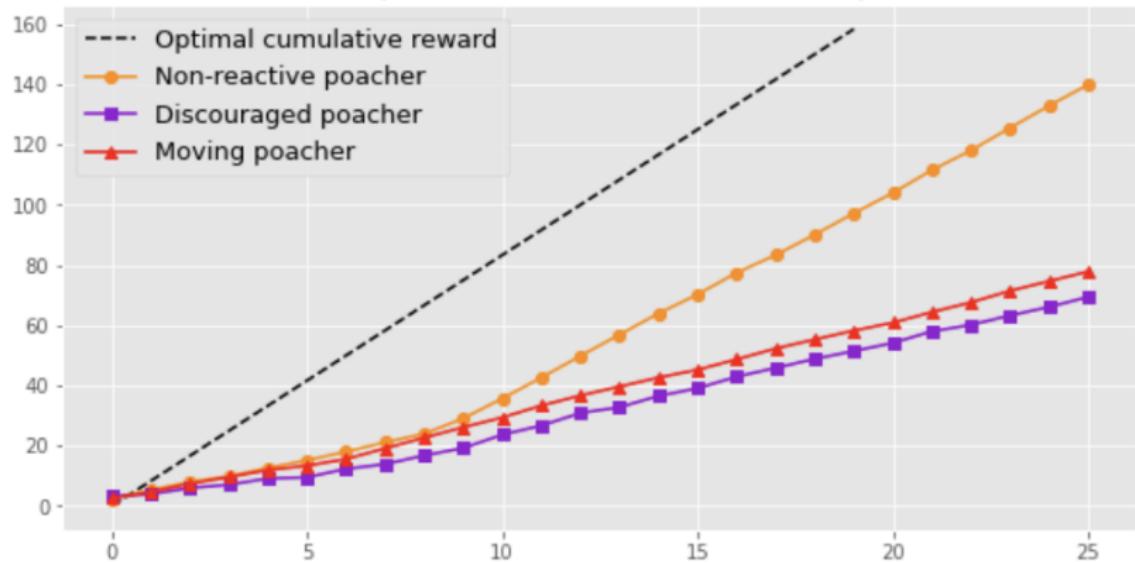
DEMO

Retornos durante 25 días:



OPONENTES REACTIVOS

Retornos durante 25 días:



Oponentes reactivos:

Como muestra la gráfica de la página anterior, el algoritmo propuesto se comporta peor respecto a oponentes reactivos.

Oponentes reactivos:

Como muestra la gráfica de la página anterior, el algoritmo propuesto se comporta peor respecto a oponentes reactivos.

Hay dos caminos naturales con los cuales seguir:

- Juegos de Stackelberg (Von Stackelberg 1934, Tambe 2012). Transformamos el problema del cazador en uno de dos etapas:

$$\min_{\beta \in C} \left(\max_{\zeta \in C} U_C(\beta, \zeta) \right)$$

Oponentes reactivos:

Como muestra la gráfica de la página anterior, el algoritmo propuesto se comporta peor respecto a oponentes reactivos.

Hay dos caminos naturales con los cuales seguir:

- Juegos de Stackelberg (Von Stackelberg 1934, Tambe 2012). Transformamos el problema del cazador en uno de dos etapas:

$$\min_{\beta \in C} \left(\max_{\zeta \in C} U_C(\beta, \zeta) \right)$$

En palabras el guardabosques se resigna a la ventaja estratégica de un cazador sofisticado e intenta sólamente limitarlo lo más posible (Green security games)

Oponentes reactivos:

Como muestra la gráfica de la página anterior, el algoritmo propuesto se comporta peor respecto a oponentes reactivos.

Hay dos caminos naturales con los cuales seguir:

- Juegos de Stackelberg (Von Stackelberg 1934, Tambe 2012). Transformamos el problema del cazador en uno de dos etapas:

$$\min_{\beta \in C} \left(\max_{\zeta \in C} U_C(\beta, \zeta) \right)$$

En palabras el guardabosques se resigna a la ventaja estratégica de un cazador sofisticado e intenta sólamente limitarlo lo más posible (Green security games) Construimos (Betancourt, -) una formulación big-M del problema resultante (como una instancia enorme de programación entera mixta). Este acercamiento sufre de problemas de escalabilidad y tiene hipótesis muy fuertes sobre los oponentes.

Oponentes reactivos:

- Hay una teoría bien desarrollada de **CMBAs adversarios** (Cesa-Bianchi y Lugosi 2012).

Oponentes reactivos:

- Hay una teoría bien desarrollada de **CMBAs adversarios** (Cesa-Bianchi y Lugosi 2012).

Fijamos un horizonte de T pasos. En cada turno t :

- ① *Los cazadores eligen actividades ilegales a realizar en un conjunto de aristas cuya función característica es el vector $\ell_t \in \{0, 1\}^m$.*
- ② *Los guardabosques eligen un subconjunto S de ciclos y patrullan (anotando la información de en qué aristas vieron actividad ilegal).*

El objetivo de los guardabosques es controlar el **regret adversario**, dado por

$$R(T) := \left[\max_{c \in \mathcal{C}} \sum_{t=1}^T \langle \ell_t, c \rangle \right] - \sum_{t=1}^T \langle \ell_t, \mathbb{E}_{q_t}[S_t] \rangle$$

Teorema. (Cesa-Bianchi y Lugosi, 2012)

Para varios CMABs existen algoritmos que cumplen

$$R(T) < \sqrt{nm \log |\mathcal{C}|}$$

Teorema. (Cesa-Bianchi y Lugosi, 2012)

Para varios CMABs existen algoritmos que cumplen

$$R(T) < \sqrt{nm \log |\mathcal{C}|}$$

El desarrollo de una generalización de este teorema para el problema de Jama Coaque es investigación en curso...

Meta-bandits:

Creemos que el algoritmo final debería ser una comunidad de algoritmos, es decir una combinación de CMAB estocástico, CMAB adversario y Stackelberg (reflejando la multiplicidad de objetivos y métodos de los posibles atacantes).

Meta-bandits:

Creemos que el algoritmo final debería ser una comunidad de algoritmos, es decir una combinación de CMAB estocástico, CMAB adversario y Stackelberg (reflejando la multiplicidad de objetivos y métodos de los posibles atacantes).

Los pesos de esta combinación deberán ser a su vez determinados mediante MABs.