Assignment 1 (2024)

Recommending News Articles to Unknown Users

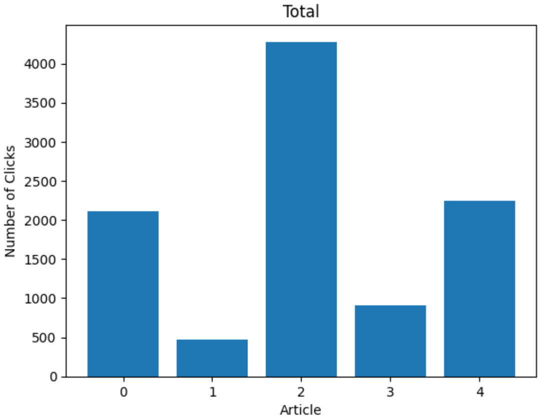
Christos Kostadimas

2020030050

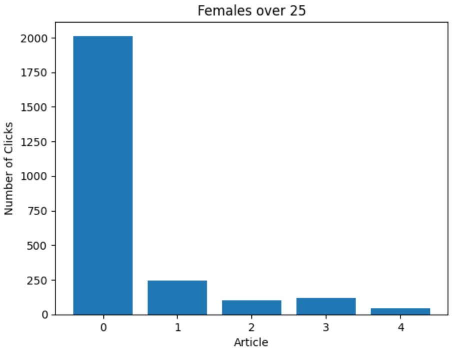
April 2024 , Chania

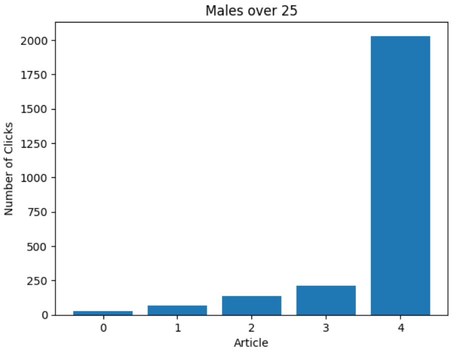
1 Plots from the simulation

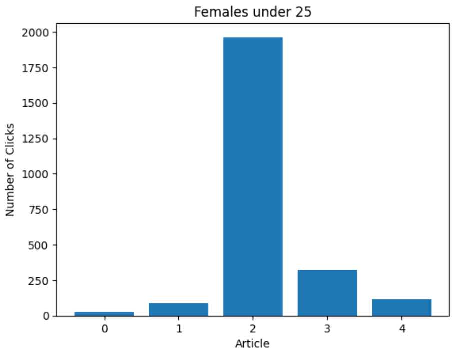
1.1 Histograms that show my algorithm truly picks the best arm for every user category:

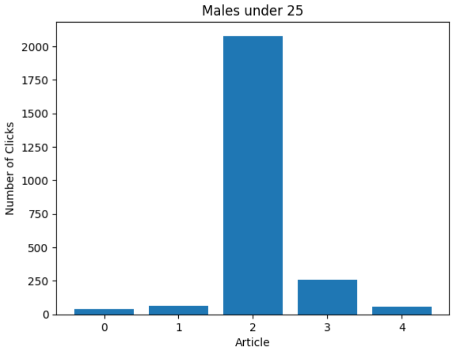


This plot is a histogram that shows the number of clicks for each article and corresponds to all user types.



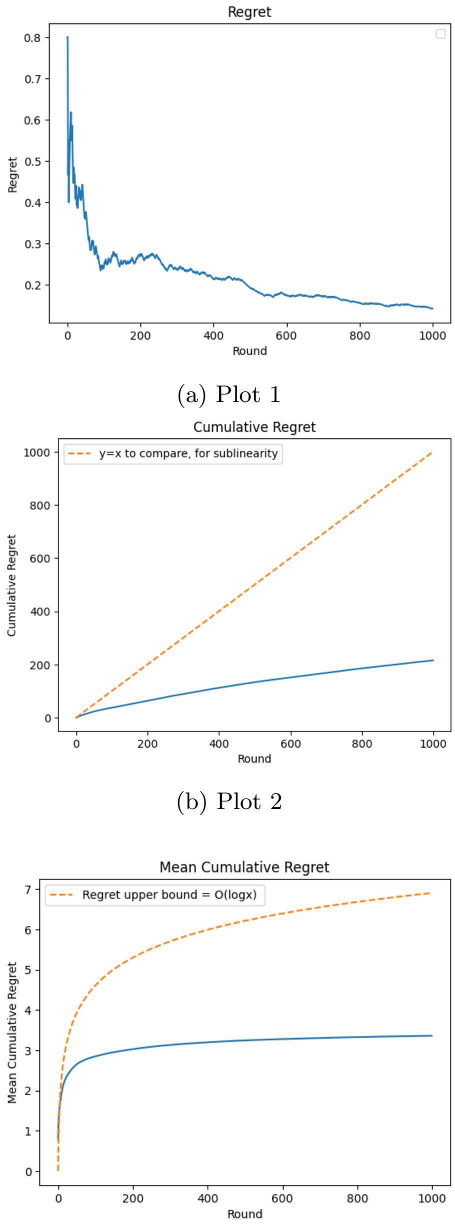



Females under 25


Males under 25

The above four plots, are histograms that show how many times my algorithm selected that specific article as the best arm, for each user type. Based on the given probabilities on the exercise , my algorithm seems to work fine.

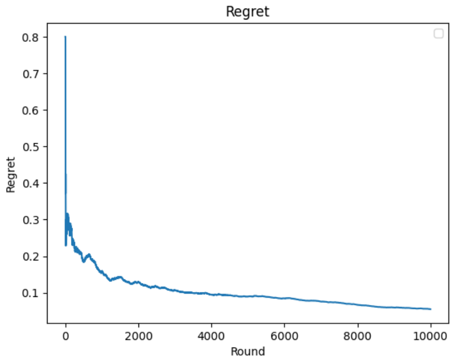
1.2 Plots for regret:



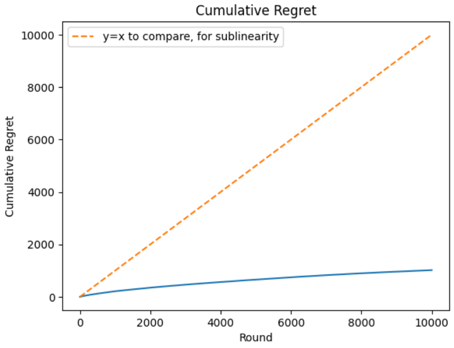
(a) Plot 1

(b) Plot 2

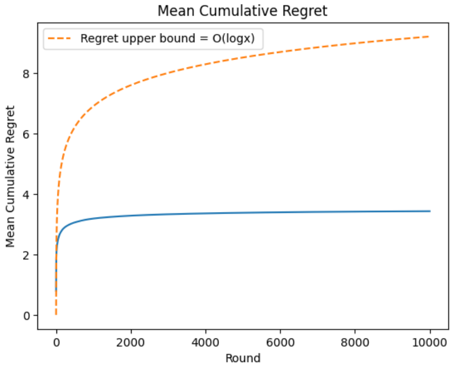
(c) Plot 3



(a) Plot 4



(b) Plot 5



(c) Plot 6

Some words about the plots:

- Plot 1 and plot 4 show the instant $regret(t)$ of round t.
- Plot 2 and plot 5 show the cumulative regret, compared to straight line $y = x$ to see if the regret is sublinear. In our case , it is.
- Plot 3 and plot 6 show the mean cumulative regret that is compared with the regret upper bound (denoted as RUB later on).

The results obtained from the simulation provide valuable insights into the performance of the algorithm under different scenarios. On the first page, I present the plots generated for a horizon of 1,000 rounds, while on the second page, I showcase the plots for a horizon of 10,000 rounds.

Upon examination of the plots, it becomes evident that increasing the number of rounds leads to a more accurate estimation of the best articles to recommend to each user type... This observation aligns with the intuition that rounds allow for a better understanding of user preferences.

Of particular interest are Plot 3 and Plot 6, which depict the mean cumulative regret compared with its upper bound, denoted as: $RUB = \log(T)$, where T represents the number of rounds.

For Plot 3, corresponding to $T = 1,000$, we observe that the mean cumulative regret $E[R_t]$ remains consistently lower than the UCB throughout the simulation.

Similarly, in Plot 6, which corresponds to $T = 10,000$, the gap between $E[R_t]$ and the $RUB = \log(T)$ (red dotted line) is notably bigger compared to Plot 3. This phenomenon underscores the effectiveness of the UCB algorithm in minimizing regret over a larger number of rounds. Formally, we can infer that as the number of rounds increases, the UCB algorithm demonstrates enhanced performance, resulting in significantly reduced total regret. This observation validates the theoretical expectation that increased exploration of the action space leads to more informed decision-making and improved outcomes.

The mean cumulative regret for each horizon (T = 1000 or T = 10000) falls below the theoretical upper bound established in Section 2

2 Upper bound proof

(Handwritten because of time limitations...)

NOTATIONS

T: Horizon
t: round

$i=0 \rightarrow$ 1st article (arm)
$\vdots$
$i=3 \rightarrow$ 4th article (arm)

$U=0 \rightarrow$ fem > 25
$U=1 \rightarrow$ male > 25
$U=2 \rightarrow$ fem < 25
$U=3 \rightarrow$ male < 25

This proof is based on the slides that exist on lecture 2 module on eClass

PROOF

In this project the environment is instance dependent.

1 Regret depends on how many times each arm (article) i is played. That number of times depends on the user type too, as I earlier showed in the histograms $\Rightarrow \boxed{N_{i,u}(t)}$

$\downarrow$ i.e. $N_{0,0}(0)$: number of times article 0 (1st) was "pulled" by a user that is a female over 25 years old

2 let: $\Delta_{i,u} = |\hat{\mu}_{i,u}(t) - \mu_{i,u}|$, $\forall u, \forall i, \forall t$

3 For a GOOD event assume:

$$\Delta_{i,u} = |\hat{\mu}_{i,u}(t) - \mu_{i,u}| \leq \sqrt{\frac{2\cdot\log(T)}{N_{i,u}(t)}} \quad \forall u \ \forall i \ \forall t$$

4 For a BAD event assume:

$$P(BAD) = P\left(\exists i,t,u : \left|\hat{\mu}_{i,u}(t) - \mu_{i,u}\right| > \sqrt{\frac{2\cdot\log(T)}{N_{i,u}(t)}}\right) \leqslant k\cdot T\cdot U\cdot T^{-4}$$

5 Assume a user belonging on user type u, pulls arm i:

- $$\left.\Delta_{i,u} \leqslant 2\cdot\sqrt{\frac{2\log(T)}{N_i(t)}}\ \right\} \Rightarrow N_{i,u}(t) \leqslant \frac{8\log T}{\Delta_{i,u}^2}$$

$$\Rightarrow \Delta_{i,u}\cdot N_{i,u}(t) \leqslant 8\log(T)\cdot\frac{1}{\Delta_{i,u}}$$

$$\Rightarrow \sum_{u=1}^{U}\sum_{i=1}^{K} \Delta_{i,u}\cdot N_{i,u}(t) \leqslant 8\log(T)\cdot\sum_{u=1}^{U}\sum_{i=1}^{K}\frac{1}{\Delta_{i,u}}$$

6 For $\mathcal{E}[R(T)]$ I have that:

$$\mathcal{E}[R(T)] = P(GOOD) \cdot \sum_{u=1}^{U} \sum_{i=1}^{K} N_{i,u}(t) + P(BAD) \cdot \sum_{u=1}^{U} \sum_{i=1}^{K} N_{i,u}(t)$$

But • $P(BAD) \leqslant KUT^{-3} \xrightarrow{\boxed{\text{if } T \to \infty}} \boxed{P(BAD) \approx 0}$

• $N_{i,u}(t) \leqslant T$ (Bad event → pulling bad arm for entire duration → max regret)

• $P(GOOD) \approx 1$

Combining the above, I get:

$$\left.\begin{aligned} \mathcal{E}[R(T)] &\leqslant \sum_{u=1}^{U} \sum_{i=1}^{K} \Delta_{i,u} \cdot N_{i,u}(t) \\ \text{Since: } \quad N_{i,u}(t) &\leqslant \frac{8 \log(t)}{\Delta_{i,u}^2} \end{aligned}\right\} \Rightarrow \mathcal{E}[R(T)] \leqslant \sum_{u=1}^{U} \sum_{i=1}^{K} \frac{8 \log T}{\Delta_{i\mu}}$$

Finally: $\mathcal{E}[R(T)] \leqslant \boxed{8 \log(t) \cdot \sum_{u=1}^{U} \sum_{i=1}^{K} \frac{1}{\Delta_{i,u}}}$

$\underline{O(\log T)}$ → seems correct for UCB instance dependent envir. (lecture 2...)

REGRET UPPER BOUND.