Showing Relevant Ads via Context Multi-Armed Bandits

Dávid Pál

December 17, 2008
A&C Seminar

joint work with Tyler Lu and Martin Pál

The Problem

- we’re running a popular website
- users visit our website
- we want to show each user relevant ad for him/her
    - relevant = likely to click on
- for each user there is some side information
    - (search query, geographic location, cookies, etc.)

Multi-Armed Bandits

- pulling an arm = showing an ad
- reward = click on the ad

Previous Work

Context-Free Multi-Armed Bandits

- historical papers by Robbins in early 1950's
- stochastic version: Lai & Robbins 1985, Auer et al. 2002
- non-stochastic version: Auer et al. 1995
- Lipschitz version: R. Kleinberg 2005, Auer et al. 2007, R. Kleinberg et al. 2008

Overview

- Our model with *context* and *Lipschitz* condition
- Regret and No-Regret learning
- Statement of our results:
    - upper and lower bound on the regret
- Our algorithm
- Idea of the analysis of the algorithm

Lipschitz Context Multi-Armed Bandits

- information x about the user (*context*)
- suppose we show ad y
- with probability $\mu(x, y)$ the user's clicks on the ad
- assume $\mu : X \times Y \to [0, 1]$ is Lipschitz:

$$|\mu(x, y) - \mu(x', y')| \leq L_X(x, x') + L_Y(y, y')$$

where L_X and L_Y are metrics

The Game

- adversary chooses $\mu : X \times Y \to [0, 1]$ and a sequence $x_1, x_2, \ldots, x_T$
- algorithm chooses $y_1, y_2, \ldots, y_T$ online:
- in round $t = 1, 2, \ldots, T$ the algorithm has access to
  - $x_1, x_2, \ldots, x_{t-1}$
  - $y_1, y_2, \ldots, y_{t-1}$
  - $\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_{t-1} \in \{0, 1\}$
- adversary reveals x_t
- based on this the algorithm outputs y_t

Regret

- optimal strategy: in round $t = 1, 2, \ldots, T$ show

$$y_t^* = \operatorname*{argmax}_{y \in Y} \ \mu(x_t, y)$$

- the algorithm shows instead $y_1, y_2, \ldots, y_T$
- difference between expected payoffs

$$\text{Regret}(T) = \sum_{t=1}^{T} \mu(x_t, y_t^*) - \mathbf{E}\left[\sum_{t=1}^{T} \mu(x_t, y_t)\right]$$

No Regret Learning

- per-round regret vanishes:

$$\lim_{T \to \infty} \frac{\text{Regret}(T)}{T} = 0$$

- how fast is the convergence? typical result:

$$\text{Regret}(T) = O(T^\gamma)$$

where $0 < \gamma < 1$.

Our Results

(Oversimplifying and lying somewhat.)

Theorem

If X has "dimension" a and Y has "dimension" b, then

- *there exists an algorithm with*

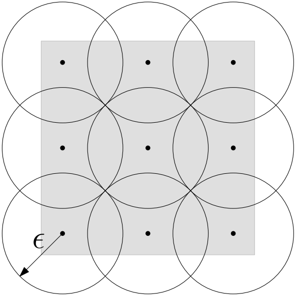
$$\text{Regret}(T) = \widetilde{O}\left(T^{\frac{a+b+1}{a+b+2}}\right)$$

- *for any algorithm*

$$\text{Regret}(T) = \Omega\left(T^{\frac{a+b+1}{a+b+2}}\right)$$

Covering Dimension

- let (Z, L_Z) be a metric space
- cover the space with ϵ-balls
- How many balls do we need?
- roughly $(1/\epsilon)^d$
- define d to be the dimension



Optimal Algorithm

- suppose that T is known to the algorithm
- X, Y have dimensions a, b respectively
- discretize X and Y:

$$\epsilon = T^{-\frac{1}{a+b+2}}$$

- X_0 are centers of ϵ-balls covering X
- Y_0 are centers of ϵ-balls covering Y
- round x_t to nearest element of X_0
- display only ads from Y_0

Optimal Algorithm, continued

- for each $x_0 \in X_0$ and $y_0 \in Y_0$ maintain:
  - number of times y_0 was displayed for x_0:

$$n(x_0, y_0)$$

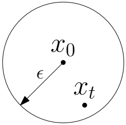
  - corresponding number of clicks:

$$m(x_0, y_0)$$

  - estimate of the click-through rate:

$$\overline{\mu}(x_0, y_0) = \frac{m(x_0, y_0)}{n(x_0, y_0)}$$

Optimal Algorithm, continued



- when x_t arrives "round" it to $x_0 \in X_0$
- show ad $y_0 \in Y_0$ that maximizes

$$\overline{\mu}(x_0, y_0) + \sqrt{\frac{\log T}{1 + n(x_0, y_0)}}$$

(exploration vs. exploitation trade-off)

Idea of Analysis

- let

$$R_t(x_0, y_0) = \sqrt{\frac{\log T}{1 + n(x_0, y_0)}}$$
$$I_t(x_0, y_0) = \overline{\mu}(x_0, y_0) + R_t(x_0, y_0)$$

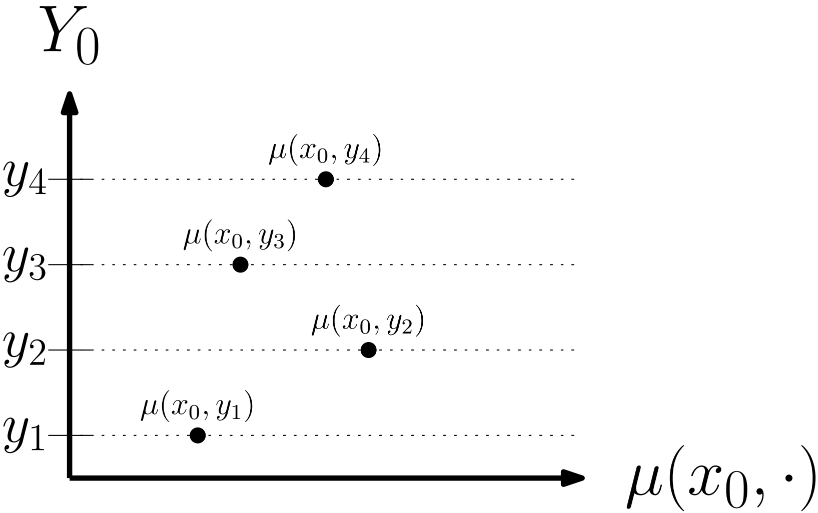
- By Chernoff-Hoeffding bound with high probability

$$I_t(x_0, y_0) \in [\mu(x_0, y_0) - \epsilon,\ \mu(x_0, y_0) + 2R_t(x_0, y_0) + \epsilon]$$

for all $x_0 \in X_0, y_0 \in Y_0$ and all $t = 1, 2, \ldots, T$ simultaneously.

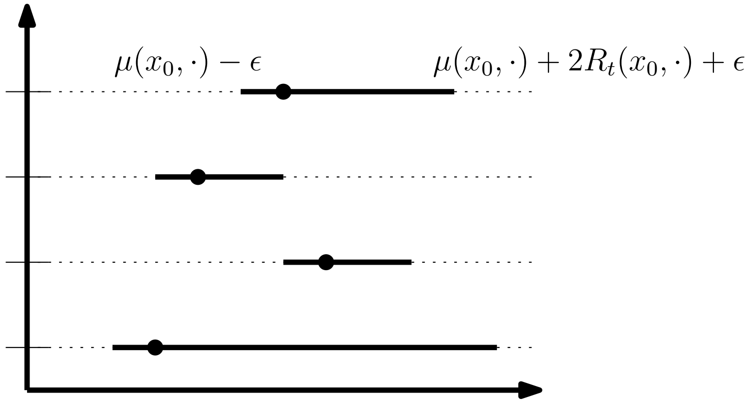
Idea of Analysis

Fix $x_0 \in X_0$



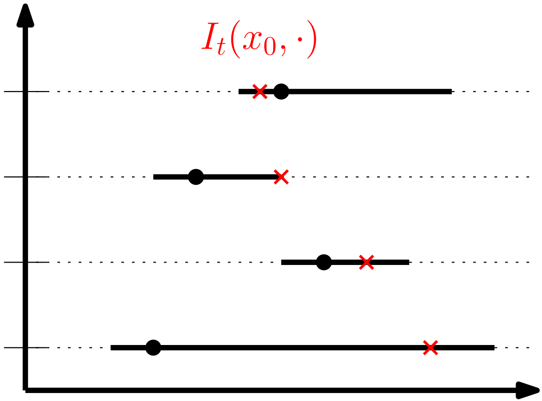
Idea of Analysis

The confidence intervals



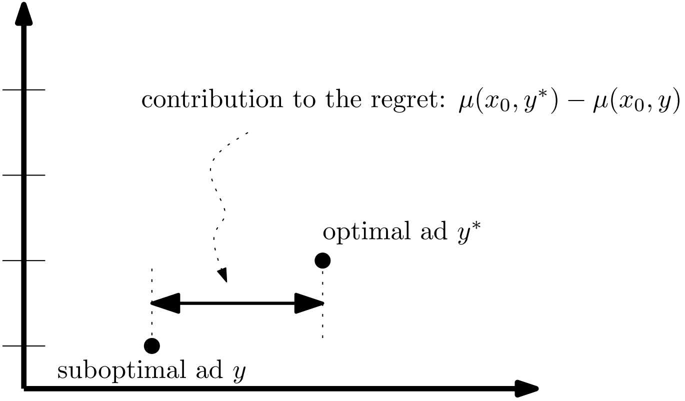
Idea of Analysis

- The algorithm displays the ad maximizing $I_t(x_0, \cdot)$.
- $I_t(x_0, y_0)$'s lies w.h.p. in the confidence interval.



Idea of Analysis

$$\text{Regret}(T) = \sum_{t=1}^{T} \mu(x_t, y_t^*) - \mathbf{E}\left[\sum_{t=1}^{T} \mu(x_t, y_t)\right]$$

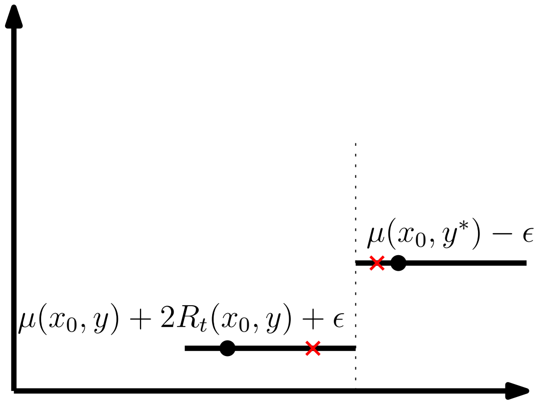


Idea of Analysis

If

$$\mu(x_0, y) + R_t(x_0, y) + \epsilon < \mu(x_0, y^*) - \epsilon \ ,$$

the algorithm stops displaying the suboptimal ad y.



Idea of Analysis

$$R_t(x_0, y) = \sqrt{\frac{\log T}{1 + n(x_0, y)}}$$

- Confidence interval for y shrinks as $n_t(x_0, y)$ increases.
- Thus we can upper bound $n_t(x_0, y)$ in terms of the difference
$$\mu(x_0, y^*) - \mu(x_0, y)$$
- Rest is just a long calculation.

Conclusion

- formulation of Context Multi-Armed Bandits
- roughly matching upper and lower bounds:

$$T^{\frac{a+b+1}{a+b+2}}$$

- `www.cs.uwaterloo.ca/~dpal/papers/`
- possible future work: non-stochastic clicks

Thanks!