Yandex

Yandex

Optimal Non-parametric Learning in Repeated Contextual Auctions with Strategic Buyer

Alexey Drutsa

Setup

Repeated Contextual Posted-Price Auctions

Different goods (e.g., ad spaces)

- described by d-dimensional feature vectors (contexts) from $[0,1]^d$
- are repeatedly offered for sale by a seller
- to a **single** buyer over T rounds (one good per round).

The buyer

- holds a private **fixed** valuation function $v: [0,1]^d \to [0,1]$
- used to calculate his valuation $v(x)$ for a good with context $x \in [0,1]^d$,
- v is **unknown** to the seller.

At each round $t = 1, \dots, T$,

- a feature vector x_t of the current good is observed by the seller and the buyer
- a price p_t is offered by the seller,
- and an allocation decision $a_t \in \{0,1\}$ is made by the buyer:

  $a_t = 0$, when the buyer rejects, and $a_t = 1$, when the buyer accepts.

Seller's pricing algorithm and buyer strategy

The seller applies a pricing algorithm A that sets prices $\{p_t\}_{t=1}^{T}$
in response to buyer decisions $\mathbf{a} = \{a_t\}_{t=1}^{T}$ and observed contexts $\mathbf{x} = \{x_t\}_{t=1}^{T}$.

The price p_t can depend only on

- past decisions $\{a_s\}_{s=1}^{t-1}$
- feature vectors $\{x_s\}_{s=1}^{t}$
- the horizon T

Strategic buyer

The seller announces her pricing algorithm A in advance

The buyer has some distribution (beliefs) D about future contexts.

In each round t, given the history of previous rounds,
he chooses his decision a_t s.t. it maximizes his future γ-discounted surplus:

$$\mathbb{E}_{x_s \sim D}\left[\sum_{s=t}^{T} \gamma^{s-1} a_s (v(x_s) - p_s)\right], \qquad \gamma \in (0,1]$$

The game’s workflow and knowledge structure

Seller's goal

The seller's strategic regret:

$$\mathrm{SReg}(T, A, v, \gamma, x_{1:T}, D) := \sum_{t=1}^{T} \left(v(x_t) - a_t^{\mathrm{Opt}} p_t\right)$$

We will learn the function v in a non-parametric way. For this, we will assume that it is Lipschitz (a standard requirement for non-parametric learning):

$$\mathrm{Lip}_L\left([0,1]^d\right) := \{f: [0,1]^d \rightarrow [0,1] \mid \forall x, y \in [0,1]^d |f(x) - f(y)| \leq L\|x - y\| \}$$

The seller seeks for a no-regret pricing for **worst-case** valuation function:

$$\sup_{v \in \mathrm{Lip}_L\left([0,1]^d\right), x_{1:T}, D} \mathrm{SReg}(T, A, v, \gamma, x_{1:T}, D) = o(T)$$

Optimality: the lowest possible upper bound for the regret of the form $O(f(T))$.

Background & Research question

Background

[Kleinberg et al., FOCS'2003]	Non-contextual setup ($d = 0$). Horizon-dependent optimal algorithm against **myopic** buyer ($\gamma = 0$) with **truthful** regret $\Theta(\log\log T)$.
[Amin et al., NIPS'2013]	Non-contextual setup ($d = 0$). The **strategic** setting is introduced. $\nexists$ no-regret pricing for non-discount case $\gamma = 1$.
[Drutsa, WWW'2017]	Non-contextual setup ($d = 0$). Horizon-**independent** optimal algorithm against **strategic** buyer with regret $\Theta(\log\log T)$ for $\gamma < 1$.
[Mao et al., NIPS'2018]	Our **non-parametric contextual** setup ($d > 0$). Horizon-dependent optimal algorithm against **myopic** buyer ($\gamma = 0$) with **truthful** regret $\Theta(T^{\frac{d}{d+1}})$.

Research question

The key approaches of the non-contextual optimal algorithms ([pre]PRRFES) **cannot be directly applied** to contextual algorithm of [Mao et al., NIPS'2018]

In order to search the valuation of the strategic buyer without context:

- Penalization rounds are used
- We **do not propose** prices below the ones that are earlier accepted

In the approach of [Mao et al., NIPS'2018]:

- Standard penalization does not help
- Proposed prices **can be below** the ones that are earlier accepted by the buyer

In this study, I overcome these issues and propose an optimal non-parametric algorithm for the contextual setting with strategic buyer

Novel optimal algorithm

Penalized Exploiting Lipschitz Search (PELS)

PELS has three parameters:

- the price offset $\eta \in [1, +\infty)$
- the degree of penalization $r \in \mathbb{N}$
- the exploitation rate $g: \mathbb{Z}_+ \to \mathbb{Z}_+$

This algorithm keeps track of

- a partition $\mathfrak{X}$ of the feature domain $[0,1]^d$
- initialized to $[(4\eta + 6)L]^d$ cubes (boxes) with side length $l = 1/[(4\eta + 6)L]$:

$$\mathfrak{X} = \{I_1 \times I_2 \times \cdots \times I_d \mid (I_1, I_2, \dots, I_d) \in \{[0, l], (l, 2l], \dots, (1 - l, 1]\}^d\}.$$

Penalized Exploiting Lipschitz Search (PELS)

For each box $X \in \mathfrak{X}$, PELS also keeps track of:

- the lower bound $u^X \in [0,1]$,
- the upper bound $w^X \in [0,1]$,
- the depth $m^X \in \mathbb{Z}_+$.

They are initialized as follows: $u^X = 0$, $w^X = 1$, and $m^X = 0$, $X \in \mathfrak{X}$.

The workflow of the algorithm is organized independently in each box $X \in \mathfrak{X}$.

- the algorithm receives a good with a feature vector $x_t \in [0,1]^d$
- finds the box $X \in \mathfrak{X}$ in the current partition $\mathfrak{X}$ s.t. $x_t \in X$.

> Then, the proposed price p_t is determined only from the current state associated with the box X, while the buyer decision a_t is used only to update the state associated with this box X.

Penalized Exploiting Lipschitz Search (PELS)

In each box $X \in \mathfrak{X}$, the algorithm iteratively offers **exploration** price:

$$u^X + \eta L\mathrm{diam}(X)$$

If this price is accepted by the buyer:

- the lower bound u^X is increased by $L\mathrm{diam}(X)$.

If this price is rejected:

- the upper bound w^X is decreased by $(w^X - u^X) - 2(\eta + 1)L\mathrm{diam}(X)$
- 1 is offered as a **penalization** price for $r - 1$ next rounds in this box X (if one of them is accepted, we continue offering 1 all the remaining rounds).

Penalized Exploiting Lipschitz Search (PELS)

- If, after an acceptance of an exploration price or after penalization rounds we have $(w^X - u^X) < (2\eta + 3)L\mathrm{diam}(X)$,
- then PELS:
  - offers the **exploitation** price u^X for $g(m^X)$ next rounds in this box X (buyer decisions made at them do not affect further pricing);
  - bisects each side of the box X to obtain 2^d boxes $\mathfrak{X}_X := \{X_1, \dots, X_{2^d}\}$ with ℓ_∞-diameter equal to $\mathbf{diam}(X)/2$;
  - refines the partition $\mathfrak{X}_X$ replacing the box X by the new boxes $\mathfrak{X}_X$.

These new boxes $\mathfrak{X}_X$

- inherit the state of the bounds u^X and w^X from the current state of X,
- while their depth $m^Y = m^X + 1 \quad \forall Y \in \mathfrak{X}_X$.

PELS is optimal

Theorem 1.

Let $d \geq 1$ and $\gamma_0 \in (0,1)$.

Then for the pricing algorithm PELS A with:

- the number of penalization rounds $r \geq \left\lceil \log_{\gamma_0} \frac{1-\gamma_0}{2} \right\rceil$
- the exploitation rate $g(m) = 2^m, m \in \mathbb{Z}_+$,
- the price offset $\eta \geq 2/(1-\gamma_0)$

for any valuation function $v \in \mathrm{Lip}_L\big([0,1]^d\big)$, discount $\gamma \leq \gamma_0$, distribution D and feature vectors $x_{1:T}$, the strategic regret is upper bounded:

$$\mathrm{SReg}(T, A, v, \gamma, x_{1:T}, D) \leq C\big(N_0(T+N_0)^d\big)^{\frac{1}{d+1}} = \Theta(T^{\frac{d}{d+1}}),$$

$$C := 2^d r(2\eta + 3 + L^{-1}) + 1 \quad \text{and} \quad N_0 := \lceil (4\eta + 6)L \rceil^d.$$

PELS: main properties and extensions

- Can be applied against myopic buyer ($\gamma = 0$) (setup of [Mao et al., NIPS'2018])
- PELS is **horizon-independent** (in contrast to [Mao et al., NIPS'2018])

What if the loss is symmetric?

- We can generalize the algorithm to classical online learning losses
- For instance, we want to optimize regret of the form $\sum_{t=1}^{T} |v(x_t) - p_t|$
- But interacting with the strategic buyer still
- Slight modification of PELS has regret $O(T^{\frac{d-1}{d}})$, which is tight for $d > 1$.

Thank you!

Alexey Drutsa

Yandex

adrutsa@yandex.ru