FEATURE AND PARAMETER SELECTION IN STOCHASTIC LINEAR BANDITS

Ahmadreza Moradipari[1], Berkay Turan[1], Yasin Abbasi-Yadkori[2]

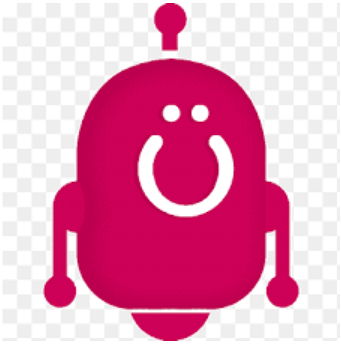
Mahnoosh Alizadeh[1], Mohammad Ghavamzadeh[3]

[1]University of California, Santa Barbara

[2] DeepMind, [3] Google Research

STOCHASTIC LINEAR BANDIT

Agent



Take Action $a_t \in \mathcal{A}_t$

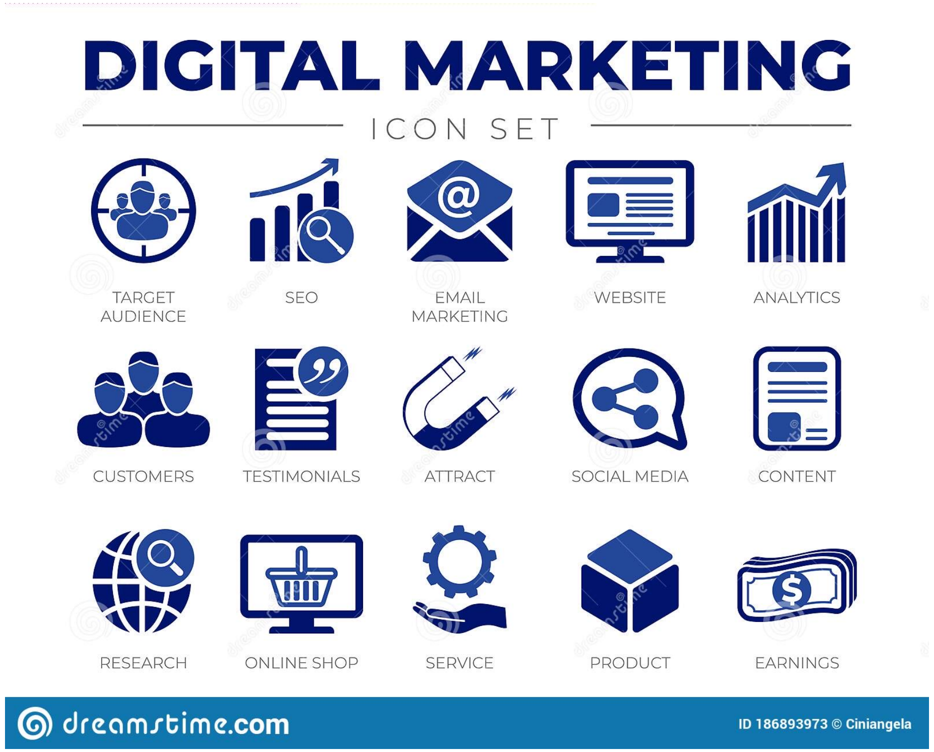
Environment

Observe reward $y_t = \langle \phi(a_t), \theta_* \rangle + \eta_t$

Regret: $\mathcal{R}(T, \theta_*) = \sum_{t=1}^{T} \langle \phi(a_t^*), \theta_* \rangle - \langle \phi(a_t), \theta_* \rangle$ **where** $a_t^* = \arg\max_{a \in \mathcal{A}_t} \langle \phi(a), \theta_* \rangle$

MODEL SELECTION IN STOCHASTIC LINEAR BANDITS

- The LB problem at hand is selected as a set of M models.
- Customers of online marketing websites belong to a certain number of categories based on their search or history of shopping



TWO MODEL SELECTION SETTINGS STUDIED

$$y_t = \langle \phi(a_t), \theta_* \rangle + \eta_t$$

TWO MODEL SELECTION SETTINGS STUDIED

$$y_t = \langle \phi(a_t), \theta_* \rangle + \eta_t$$

Feature selection

- The agent is given a set of M feature maps $\{\phi^i\}_{i=1}^M$
- Expected reward belongs to the linear span of at least one of these feature maps, i.e., $\exists i \in [M]$:

$$\mathbb{E}[y_t] = \langle \phi^i, \theta_*^i \rangle$$

- We minimize the transfer regret:

$$\mathcal{R}(T) = \sum_{t=1}^{T} \langle \phi^{i_*}(a_t^*), \theta_*^{i_*} \rangle - \langle \phi^{i_*}(a_t), \theta_*^{i_*} \rangle$$

Here i_* is the true model.

TWO MODEL SELECTION SETTINGS STUDIED

$$y_t = \langle \phi(a_t), \theta_* \rangle + \eta_t$$

Feature selection

- The agent is given a set of M feature maps $\{\phi^i\}_{i=1}^M$
- Expected reward belongs to the linear span of at least one of these feature maps, i.e., $\exists i \in [M]$:

$$\mathbb{E}[y_t] = \langle \phi^i(a_t), \theta_*^i \rangle$$

- We minimize the transfer regret:

$$\mathcal{R}(T) = \sum_{t=1}^{T} \langle \phi^{i_*}(a_t^*), \theta_*^{i_*} \rangle - \langle \phi^{i_*}(a_t), \theta_*^{i_*} \rangle$$

Here i_* is the true model.

Parameter selection

- Reward parameter can be generated from M possible reward models each defined as a ball

$$B(\mu_i, b_i) = \{\theta \in \mathbb{R}^d : \|\theta - \mu_i\| \leq b_i\}$$

- We assume estimates of the centers $\{\widehat{\mu}_i\}_{i=1}^M$ with upper bound on the errors of these estimates are available $\|\mu_i - \widehat{\mu}_i\| \leq c_i, \;\; \forall i \in [M]$
- We minimize the transfer regret:

$$\mathcal{R}(T) = \sup_{\theta_* \in \bigcup_{i=1}^M B(\mu_i, b_i)} \mathcal{R}(T, \theta_*)$$

where

$$\mathcal{R}(T, \theta_*) = \sum_{t=1}^{T} \langle \phi(a_t^*), \theta_* \rangle - \langle \phi(a_t), \theta_* \rangle$$

FS-SCB REGRET

Theorem 1. *Let Assumption 1 hold and the regularization parameters λ_i, exploration parameter κ, and learning rate α set to the values described above. Then, for any $\delta \in [0, 1/4)$, with probability at least $1 - \delta$, the regret defined by (2) for FS-SCB is bounded as*

$$\mathcal{R}_{\text{FS-SCB}}(T) \leq \mathcal{O}\left(\sqrt{2T\log(2/\delta)} + RLG\sqrt{KT(1+\log(M))\max_{i\in[M]}\left\{\lambda_i S^2 + 4d\log\left(\frac{1+\frac{TL^2}{\lambda_i d}}{\delta}\right)\right\}}\right).$$

FS-SCB REGRET

Theorem 1. *Let Assumption 1 hold and the regularization parameters λ_i, exploration parameter κ, and learning rate α set to the values described above. Then, for any $\delta \in [0, 1/4)$, with probability at least $1-\delta$, the regret defined by (2) for FS-SCB is bounded as*

$$\mathcal{R}_{\text{FS-SCB}}(T) \leq \mathcal{O}\left(\sqrt{2T\log(2/\delta)} + RLG\sqrt{KT(1+\log(M))\max_{i\in[M]}\left\{\lambda_i S^2 + 4d\log\left(\frac{1+\frac{TL^2}{\lambda_i d}}{\delta}\right)\right\}}\right).$$

Logarithmic dependence on number of the models

PS-OFUL REGRET

Theorem 6. *Let Assumption 2 hold and* $\lambda_i = \frac{1}{(b_i+c_i)^2} \geq 1, \forall i \in [M]$*. Then, for any* $\delta \in (0, 1/4]$*, with probability at least* $1-\delta$*, the transfer-regret defined by (3) of PS-OFUL is bounded as*

$$\mathcal{R}(T) = \mathcal{O}\left(dRL \max\{1, G\} + \sqrt{1+\log(M)} \sqrt{T \log\left(1 + \frac{T}{d}\right) \log\left(\frac{1 + \frac{TL^2 \max_{i\in[M]}(b_i+c_i)^2}{d}}{\delta} \right)} \right). \quad (10)$$

PS-OFUL REGRET

Theorem 6. *Let Assumption 2 hold and* $\lambda_i = \frac{1}{(b_i+c_i)^2} \geq 1, \forall i \in [M]$. *Then, for any* $\delta \in (0, 1/4]$, *with probability at least* $1-\delta$, *the transfer-regret defined by* (3) *of PS-OFUL is bounded as*

$$\mathcal{R}(T) = \mathcal{O}\left(dRL\max\{1,G\} + \sqrt{1+\log(M)}\sqrt{T\log\left(1+\frac{T}{d}\right)\log\left(\frac{1+\frac{TL^2\max_{i\in[M]}(b_i+c_i)^2}{d}}{\delta}\right)}\right). \quad (10)$$

Logarithmic dependence on number of the models

Logarithmic dependence on model misspecifications

Thank you!
Q/A?