A Simple Unified Framework for High Dimensional Bandit Problems

Wenjie Li, Adarsh Barik, Jean Honorio
li3549@purdue.edu

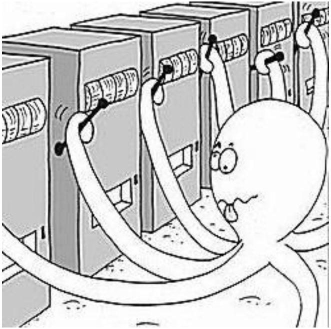
Department of Statistics
Department of Computer Science
Purdue University

July 10, 2022

Outline

Introduction and Notations

Stochastic multiarmed contextual bandits are useful models in various application domains, such as recommendation systems, online advertising, and personalized healthcare (Auer 2002b; Chu et al. 2011; Abbasi-Yadkori, Pal, and Szepesvari 2011).



Introduction and Notations (cont.)

In practice, such problems are often high-dimensional, but the unknown parameter is typically assumed to have low-dimensional structure, which in turns implies a succinct representation of the final reward.

Examples

- LASSO Bandit
- Low Rank Matrix Bandit
- Group Sparse Matrix Bandit

Introduction and Notations (cont.)

However, prior works are scattered and different algorithms with different assumptions are proposed for these problems. In this work, our contributions are

- We present a simple and unified algorithm framework named Explore-the-Structure-Then-Commit (ESTC) for high dimensional stochastic bandit problems
- We provide a problem-independent regret analysis framework for our algorithm.
- We demonstrate the usefulness of our framework by applying it to different high dimensional bandit problems.

Introduction and Notations (cont.)

A set of contexts $\{x_{t,a_i}\}_{i=1}^{K}$ for each arm is generated at every round t, and then the agent chooses an action a_t from the K arms. The contexts are assumed to be sampled i.i.d from a distribution $\mathcal{P}_X$ with respect to t, but the contexts for different arms can be correlated (Chu et al. 2011). After the action is selected, a reward $y_t = f(x_{t,a_t}, \theta^*) + \epsilon_t$ for the chosen action is received.

Let $a_t^* = \text{argmax}_{i \in [K]} f(x_{t,a_i}, \theta^*)$ denote the optimal action at each round. We measure the performance of all algorithms by the expectation of the regret, denoted as

$$\mathbb{E}[\text{Regret}(T)] = \mathbb{E}\left[\sum_{t=1}^{T} f(x_{t,a_t^*}, \theta^*) - f(x_{t,a_t}, \theta^*)\right]$$

Algorithm

Algorithm 1 Explore-the-Structure-Then-Commit (ESTC)

Input: $\lambda_{T_0}, K \in \mathbb{N}, L_t(\theta), R(\theta), f(x, \theta), \theta_0, T_0$
Initialize $\mathbf{X}_0, \mathbf{Y}_0 = (\emptyset, \emptyset), \theta_t = \theta_0$
for $t = 1$ to T_0 **do**
    Observe K contexts, $x_{t,1}, x_{t,2}, \cdots, x_{t,K}$
    Choose action a_t uniformly randomly
    Receive reward $y_t = f(x_{t,a_t}, \theta^*) + \epsilon_t$
    $\mathbf{X}_t = \mathbf{X}_{t-1} \cup \{x_{t,a_t}\}, \mathbf{Y}_t = \mathbf{Y}_{t-1} \cup \{y_{a_t}\}$
end for
Compute the estimator θ_{T_0}:

$$\theta_{T_0} \in \text{argmin}_{\theta \in \mathbf{\Theta}} \left\{ L_{T_0}(\theta; \mathbf{X}_{T_0}, \mathbf{Y}_{T_0}) + \lambda_{T_0} R(\theta) \right\}$$

for $t = T_0 + 1$ to T **do**
    Choose action $a_t = \text{argmax}_a f(x_{t,a}, \theta_{T_0})$
end for

Algorithm (cont.)

Our algorithm generalizes over the prior efforts on different high dimensional bandit problems.

Advantages

- it is very simple
- it does not require strong assumptions
- it can be applied to different problems

Theorem

(Problem Independent Regret Bound) *The expected cumulative regret of Algorithm 1 satisfies the bound*

$$\mathbb{E}[Regret(T)] = \mathcal{O}\left(\sum_{t=T_0}^{T} \sqrt{9\frac{\lambda_{T_0}^2}{\alpha^2}\phi^2 + \frac{1}{\alpha}[2Z_{T_0}(\theta^*) + 4\lambda_{T_0}R(\theta^*_{\mathcal{M}^\perp})]}\right)$$

Theoretical Results (cont.)

Table 1: Summary of Regret Bounds of Our ESTC Algorithm Framework in Different High Dimensional Bandit Problems

High Dimensional Bandit Problem	Regret Bound
LASSO Bandit (sanity check)	$\mathcal{O}(s^{1/3}T^{2/3}\sqrt{\log(dT)})$
Low-rank Matrix Bandit	$\mathcal{O}(r^{1/3}T^{2/3}\log((d_1+d_2)T)$
Group-Sparse Matrix Bandit	$\mathcal{O}(s^{1/3}\sqrt{d_2}T^{2/3}+s^{1/3}T^{2/3}\sqrt{\log d_1 T})$
Multi-agent Matrix Bandit	$\mathcal{O}(d_2 s^{1/3}T^{2/3}\sqrt{\log(d_1 T)})$

Thank you!

Thank you!

Bibliography

Abbasi-Yadkori, Yasin, David Pal, and Csaba Szepesvari (2011). "Online-to-Confidence-Set Conversions and Application to Sparse Stochastic Bandits". In: *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*.

Auer, Peter (2002b). "Using confidence bounds for exploitationexploration trade-offs". In: *Journal of Machine Learning Research*, 3:397–422.

Chu, Wei et al. (2011). "Contextual Bandits with Linear Payoff Functions". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*.