Distributionally-Aware Kernelized Bandit Problems for Risk Aversion

Sho Takemori

Fujitsu Limited

July, 2022

Introduction and Overview

In this work, we consider the following generalization of the kernelized bandit problem.

- Algorithms try to optimize risk-averse metrics (instead of the mean) such as the Mean-Variance or CVaR of the outputs of known function (or probability kernel).
- Let y be an output random variable at a point x, F the CDF of the output y. Then CVaR is defined as $\mathbf{E}\left[y \mid y \leq F^{-1}(\alpha)\right]$, where $\alpha \in (0, 1)$ is a parameter of the metric.
- However, most existing works on optimization of such risk-averse metrics have restrictions (e.g., they need the environment random variable).
- In this work, we address the issues by modeling the output distributions using kernel mean embeddings (KME) and a probability kernel.
- Then, we propose UCB-type and phased-elimination based algorithms for CVaR and MV, and prove a near optimality.

Comparison with Existing Work

- In most existing works on kernelized bandit problems for risk-aversion, they model the output y by $y = f(x, W)$, where x is an input variable, and W is a RV called the environment RV that accounts for randomness of the output y.
- However, usually, algorithms based on this model have some limitations or shortcomings.
- Recently, Nguyen et al. (2021) proposed kernelized bandit algorithms for CVaR, they assumed that algorithms can control/select W in optimization procedure, which is a restrictive assumption for complex environments (such as the real world).
- Moreover, since the regret upper bound is given using the maximum information gain of a function w.r.t. (x, W), their algorithms can have larger regret upper bounds due to possible high dimensionality of W even if that of x is moderate.

Notation and Brief Review of Kernel Mean Embeddings

- $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and $l : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ be kernels on sets $\mathcal{X}$ and $\mathcal{Y}$ with $\mathcal{Y} \subset \mathbb{R}$.
- Let $\phi_k : \mathcal{X} \to \mathcal{H}_k(\mathcal{X})$ be the feature map to the RKHS $\mathcal{H}_k(\mathcal{X})$ define ϕ_l similarly.
- Under mild conditions on the kernel l, $\exists! \mu_l : \mathcal{M}(\mathcal{Y}) \to \mathcal{H}_l(\mathcal{Y})$ s.t.

$$\langle \mu_l(\rho), f \rangle_l = \mathbf{E}_{y\sim\rho}\left[f(y)\right], \quad \forall f \in \mathcal{H}_l(\mathcal{Y}).$$

  Here $\langle \cdot, \cdot \rangle_l$ denotes the inner product in $\mathcal{H}_l(\mathcal{Y})$ and $\mathcal{M}(\mathcal{Y})$ denotes the space of probability distributions on $\mathcal{Y}$.
- The map μ_l is called Kernel Mean Embedding (KME).

Problem Formulation

- For unknown map $\boldsymbol{\rho} : \mathcal{X} \to \mathcal{M}(\mathcal{Y})$ and a given time interval T, an agent selects an arm $x_t \in \mathcal{X}$ based on the observation history $x_1, y_1, \ldots, x_{t-1}, y_{t-1}$ for each round $t = 1, \ldots, T$.
- The environment reveals a noisy output y_t with $y_t \mid \mathcal{F}_{t-1} \sim \boldsymbol{\rho}(x_t)$, where $\mathcal{F}_{t-1}$ denotes the σ-algebra generated by $x_1, y_1, \ldots, x_t$.
- The performance of an algorithm is evaluated by the cumulative CVaR regret defined as

$$R_{\mathrm{CVaR},\alpha}(T) = \sum_{t=1}^{T} \left(\sup_{x \in \mathcal{X}} \mathrm{CVaR}_\alpha(\boldsymbol{\rho}(x)) - \mathrm{CVaR}_\alpha(\boldsymbol{\rho}(x_t)) \right).$$

Model Assumption: Probability Kernel Embedding Approach

- Without smoothness assumption one cannot hope for an algorithm with a sublinear regret guarantee.
- In the commutative diagram (i.e., $\Theta \circ \phi_k = \mu_l \circ \boldsymbol{\rho}$) below, the map Θ controls the smoothness of $\boldsymbol{\rho}$.
- In this paper, we assume that Θ is a bounded linear operator between RKHSs.
- If l is the linear kernel, this model assumption is identical to the conventional model assumption in the kernelized bandit problem.
- This assumption is closely related to conditional mean embeddings, but we consider a more suitable setting for the bandit problem (e.g., initially, a probability kernel $\boldsymbol{\rho}$ is given).

$$
\begin{array}{ccc}
\mathcal{X} & \xrightarrow{\boldsymbol{\rho}} & \mathcal{M}(\mathcal{Y}) \\
\text{feature map } \phi_k \downarrow & & \text{KME } \mu_l \downarrow \\
\mathcal{H}_k(\mathcal{X}) & \xrightarrow{\Theta} & \mathcal{H}_l(\mathcal{Y})
\end{array}
$$

A UCB-type Algorithm

For observation history $(x_1, y_1), \ldots, (x_t, y_t)$ up to time step t, we define $\widehat{\mathrm{CVaR}}_{\alpha,t}(x)$ by

$$\sup_{\nu \in \mathcal{Y}} \left\{ \nu - \frac{1}{\alpha}(\psi_\nu(y_1), \ldots, \psi_\nu(y_t))(\mathbf{k}(x_{1:t}, x_{1:t}) + \lambda 1_t)^{-1}\mathbf{k}(x_{1:t}, x) \right\}, \tag{1}$$

where $\mathbf{k}(x_{1:t}, x_{1:t}) = (k(x_i, x_j))_{1 \le i,j \le t}$, $\mathbf{k}(x_{1:t}, x)^{\mathrm{T}} = (k(x_i, x))_{1 \le i \le t}$, and $\psi_\nu(y) = \max\{\nu - y, 0\}$. Assuming $|\mathcal{Y}| < \infty$, with probability at least $1 - \delta$, we have

$$\left| \mathrm{CVaR}_\alpha(\boldsymbol{\rho}(x)) - \widehat{\mathrm{CVaR}}_{\alpha,t}(x) \right| \le \frac{U}{\alpha}\beta_{k,t}^{(\mathrm{CV})}(\delta)\sigma_{k,t}(x), \tag{2}$$

for all x and t, where $\beta_{k,t}^{(\mathrm{CV})}(\delta) = O(\sqrt{(\gamma_{k,t} + \log(|\mathcal{Y}|/\delta))})$ and $\gamma_{k,t}$ is the maximum information gain.

Theorem

We can consider a UCB-type algorithm for CVaR, and with probability at least $1 - \delta$ its cumulative regret is upper bounded by $O(\frac{1}{\alpha}\beta_{k,t}^{(\mathrm{CV})}(\delta)\sqrt{T\gamma_{k,T}})$.

Rough Statement for a Nearly Optimal Algorithm

We can consider a phased algorithm (as in the conventional setting) for CVaR and provide a rough statement of the results.

Theorem

- *Assume that $\mathcal{X}$ and $\mathcal{Y}$ are finite. Then, with probability at least $1-\delta$, the cumulative regret of the phased algorithm is upper bounded by $\widetilde{O}(\frac{1}{\alpha}\sqrt{log(|\mathcal{X}||\mathcal{Y}|/\delta)}\sqrt{T\gamma_{k,T}})$.*
- *Moreover, if k is a Matérn kernel, then the phased algorithm is nearly optimal, i.e., up to a poly-logarithmic factor of T, the upper bound matches an algorithm-independent lower bound of the problem.*

Experiments in Synthetic Environments

- We empirically compare the UCB-type algorithm for CVaR and IGP-UCB in the case when $\mathcal{X}$ is a discretization of $[0, 1]^3$.
- We randomly generate lognormal environments $\mathcal{LN}(\mu_m(x), \sigma_m(x))$ by randomly generated functions $\mu_m(x), \sigma_m(x)$ for $m = 1, \dots, 10$.
- As the theoretical result indicates the proposed method incurs sublinear regret for all α and outperforms the baseline algorithm in many cases.

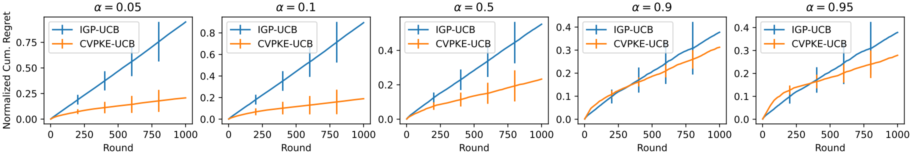

Figure: Cumulative CVaR Regret for LogNormal Environments

References I

Quoc Phong Nguyen, Zhongxiang Dai, Bryan Kian Hsiang Low, and Patrick Jaillet. Optimizing conditional Value-At-Risk of black-box functions. *Advances in Neural Information Processing Systems*, 34, 2021.