Approximation Theory Based Methods for RKHS Bandits

Sho Takemori and Masahiro Sato

FUJIFILM Business Innovation

ICML 2021

Introduction

- To tackle existing issues in the RKHS bandit problem (or kernelized bandit problem), we propose a novel reduction method from RKHS bandit problems to (misspecified) linear bandit problems.
- Specifically, we address the following issues using the reduction method:
  - Non-existence of algorithms for the adversarial RKHS bandit problem with general reward functions.
  - High computational complexity of the stochastic RKHS bandit algorithms.
- We derive the reduction method from an approximation method (P-greedy) developed in the approximation theory literature and it could potentially solve issues beyond the above.

Function Approximation in RKHS

Let $K : \Omega \times \Omega \to \mathbb{R}$ be a positive definite kernel and $\mathcal{H}_K(\Omega)$ the corresponding RKHS, where $\Omega \subset \mathbb{R}^d$ is a subset.

- Usual approximation methods used for the RKHS bandit problem basically aim to approximate the value of kernel $K(x, y)$ by the inner product of finite dimensional vectors.
- In this talk, we consider approximation of functions in the RKHS by an element of a finite dimensional subspace of the RKHS.

Reduction from RKHS Bandits to Misspecified Linear Bandits

To approximate a function in the RKHS, we apply a greedy algorithm called the P-greedy.

The P-greedy algorithm takes an admissible error $\mathfrak{e}$ (or tolerance) and returns a finite number of functions (called Newton basis) $N_1, \ldots, N_D$.

Then for any $f \in \mathcal{H}_K(\Omega)$ and x, we have

$$|f(x) - \underbrace{\langle \theta_f, \widetilde{x} \rangle}_{\text{linear model}} | \leq \underbrace{\|f\|_{\mathcal{H}_K(\Omega)} \mathfrak{e}}_{\text{misspecification error}},$$

where $\theta_f = \left(\langle f, N_i \rangle_{\mathcal{H}_K(\Omega)}\right)_{1 \leq i \leq D} \in \mathbb{R}^D$ and $\widetilde{x} = (N_i(x))_{1 \leq i \leq D} \in \mathbb{R}^D$.

Reduction from RKHS Bandits to Misspecified Linear Bandits

- If f is a reward function of a RKHS bandit problem, we can regard the problem as a misspecified linear bandit problem.
- We can construct algorithms for a misspecified linear bandit problem by modifying existing ones for linear bandits.

Convergence Rate of the P-greedy algorithm

Theorem 1 (Santin and Haasdonk 2017)

Let $\alpha, q, T > 0$ and denote by $D = D_{q,\alpha}(T)$ the number of functions returned by the P-greedy algorithm with error $\epsilon = \alpha / T^q$.

1. *Suppose K has finite smoothness with parameter $\nu > 0$. Then $D_{q,\alpha}(T) = O\left(\alpha^{-d/\nu} T^{dq/\nu}\right)$.*
2. *Suppose K has infinite smoothness. Then $D_{q,\alpha}(T) = O\left((q \log T - \log(\alpha))^d\right)$.*

We omit the definition of smoothness of kernels (see the paper for the definition). We note that rational quadratic and squared exponential kernels have infinite smoothness and Matern kernels with parameter ν have finite smoothness.

Main Results (Stochastic Case, APG-UCB)

We apply a modification of LinUCB to the stochastic RKHS bandit problem and call the algorithm APG-UCB. (Here APG stands for Approximation theory based method using P-Greedy.)

Theorem 2

Let $R_{\text{APG-UCB}}(T)$ be the (cumulative) regret that APG-UCB incurs for the stochastic RKHS bandit problem up to time step T. Then with probability at least $1-\delta$, $R_{\text{APG-UCB}}(T)$ is given as

$$\widetilde{O}\left(\sqrt{TD_{q,\alpha}(T)\log(1/\delta)} + D_{q,\alpha}(T)\sqrt{T}\right)$$

and the total computational complexity of the algorithm is given as $O(|\mathcal{A}|TD_{q,\alpha}^2(T))$.

In the paper, we also showed that APG-UCB is an approximation of IGP-UCB (Chowdhury and Gopalan 2017), whose total computational complexity is given as $O(|\mathcal{A}|T^3)$.

Main Results (Adversarial Case)

Next, we apply EXP3 for adversarial linear bandits to the adversarial RKHS bandit problem.

Theorem 3

Let $R_{APG\text{-}EXP3}(T)$ be the cumulative regret that APG-EXP3 with $\alpha = \log(|\mathcal{A}|)$ and $q = 1$ incurs for the adversarial RKHS bandit problem up to time step T. Then the expected regret $\mathrm{E}\left[R_{APG\text{-}EXP3}(T)\right]$ is given as $\widetilde{O}\left(\sqrt{TD_{1,\alpha}(T)\log\left(|\mathcal{A}|\right)}\right)$.

Remark 4

Chatterji et al. 2019 also proved a similar result. However, they only consider "the kernel loss case" (i.e., the case when the objective function f_t has a form $K(\cdot, \xi)$, which is a very special function in the RKHS).

Experiments in Synthetic Environments

Although APG-UCB has empirically almost the same cumulative regret as IGP-UCB, its running time is much shorter than IGP-UCB, which supports our theoretical results.

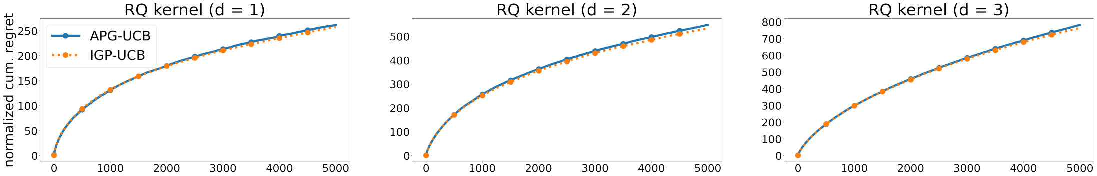

Figure: Normalized Cumulative Regret for RQ kernels.

Table: Total Running Time (in seconds).

	APG-UCB	IGP-UCB
$d = 1$	4.2e-01	5.7e+03
$d = 2$	2.7e+00	5.1e+03
$d = 3$	3.0e+01	5.7e+03

References I

Chatterji, Niladri, Aldo Pacchiano, and Peter Bartlett (2019). "Online learning with kernel losses". In: *International Conference on Machine Learning*. PMLR, pp. 971–980.

Chowdhury, Sayak Ray and Aditya Gopalan (2017). "On kernelized multi-armed bandits". In: *Proceedings of the 34th International Conference on Machine Learning*, pp. 844–853.

Santin, Gabriele and Bernard Haasdonk (2017). "Convergence rate of the data-independent *P*-greedy algorithm in kernel-based approximation". In: *Dolomites Research Notes on Approximation* 10.Special_Issue.