Open Problem: Tight Online Confidence Intervals for RKHS Elements

Sattar Vakili, MediaTek Research, UK

Jonathan Scarlett, National University of Singapore

Tara Javidi, University of California, San Diego

COLT, August 19th, 2021

Introduction

- ◇ Kernel: an elegant technique to extend linear models to non-linear
    - ◇ GP-UCB [Srinivas et al., 2010], GP-TS [Chowdhury and Gopalan, 2017], GP-EI [Nguyen et al., 2017]
    - ◇ NeuralUCB [Zhou et al., 2020], NeuralTS [Zhang et al., 2021],
    - ◇ KOVI [Yang et al., 2020]
- ◇ Suboptimal regret bounds
- ◇ Likely reason: loose confidence intervals
- ◇ Main challenge: online nature of the observation points

Kernelized Bandit

◇ An online learning algorithm collects a sequence of noisy observations $\{(x_n, y_n)\}_{n=1}^{\infty}$

$$y_n = f(x_n) + \epsilon_n$$

◇ Performance measure:

$$\mathcal{R}(N) = \sum_{n=1}^{N} (f(x^*) - f(x_n))$$

◇ Well-behaved noise: R sub-Gaussian

◇ Assumption: The RKHS norm of f is bounded

$$\|f\|_{\mathcal{H}_k} \leq B.$$

Kernel-Based Models

◇ A positive definite kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

◇ Reproducing kernel Hilbert space (RKHS) $\mathcal{H}_k$:

$$f \in \mathcal{H}_k \iff f(x) = \sum_{m=1}^{\infty} w_m \lambda_m^{\frac{1}{2}} \phi_m(x)$$

$$\|f\|_{\mathcal{H}_k} = \|\mathbf{w}\|_{l^2}$$

Surrogate Gaussian Process Model

◇ A surrogate GP model F

◇ Prediction: $\mu_n(x) = \mathbb{E}\big[F(x) | \{(x_i, y_i)\}_{i=1}^n\big]$

◇ Uncertainty estimate: $\sigma_n^2(x) = \mathbb{E}\big[(F(x) - \mu_n(x))^2 | \{(x_i, y_i)\}_{i=1}^n\big]$

$$\mu_n(x) = \mathbf{k}_n^\top(x)(\lambda^2 \mathbf{I}_n + \mathbf{K}_n)^{-1}\mathbf{y}_n$$

$$\sigma_n^2(x) = k(x, x) - \mathbf{k}_n^\top(x)(\lambda^2 \mathbf{I}_n + \mathbf{K}_n)^{-1}\mathbf{k}_n(x),$$

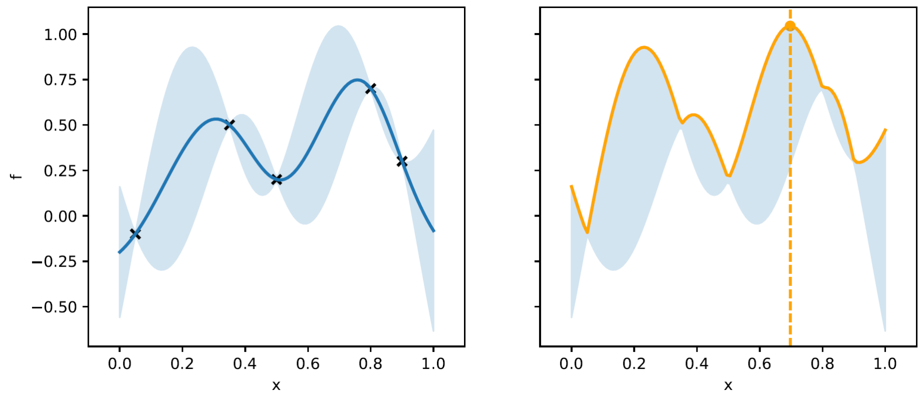
- $\mathbf{k}_n(x) = [k(x, x_1), k(x, x_2), \ldots, k(x, x_n)]^\top$
- $[\mathbf{K}_n]_{i,j} = k(x_i, x_j)$

Kernelized Bandit aka GP Bandit, Bayesian Optimization,

Algorithm

◇ GP-UCB:

$$x_n = \mathrm{argmax}_{x \in \mathcal{X}} \ \mu_{n-1} + \rho_n(\delta)\sigma_{n-1}(x)$$



$$\mathcal{R}(N) = \tilde{\mathcal{O}}(\rho_N(\delta)\sqrt{N\gamma_N}), \ \text{w.p. } 1-\delta$$

$$\gamma_n = \sup_{\{\mathbf{x}_i\}_{i=1}^n \subset \mathcal{X}} \log\det(\mathbf{I}_n + \frac{\mathbf{K}_n}{\lambda^2})$$

Online Confidence Intervals for RKHS Elements

Theorem [Abbasi-Yadkori, 2013, Chowdhury and Gopalan, 2017]. When $\|f\|_{\mathcal{H}_k} \leq B$, in the online setting, with probability $1-\delta$, for all $x \in \mathcal{X}$

$$|f(x) - \mu_n(x)| \leq \rho_n(\delta)\sigma_n(x)$$

- $\rho_n(\delta) = B + R\sqrt{2\left(\boldsymbol{\gamma_{n-1}} + 1 + \log(\frac{1}{\delta})\right)}$
- $\gamma_n = \sup_{\{\mathbf{x}_i\}_{i=1}^n \subset \mathcal{X}} \log\det(\mathbf{I}_n + \frac{\mathbf{K}_n}{\lambda^2})$
- $\gamma_n \sim D_{\text{eff},n}$

Open Problem

Open Problem: When $\|f\|_{\mathcal{H}_k} \leq B$, in the online setting, consider confidence interval

$$|f(x) - \mu_n(x)| \leq \rho_n(\delta)\sigma_n(x) \text{ w.p. } 1-\delta$$

◇ What is the lowest growth rate of $\rho_n(\delta)$ with n?

◇ Is it possible to reduce the confidence interval width by an $\tilde{\mathcal{O}}(\sqrt{\gamma_n})$ factor?

Regret Bounds

◇ Regret Bound

$$\mathcal{R}(N) = \tilde{\mathcal{O}}(\rho_N(\delta)\sqrt{N\gamma_N}), \text{ w.p. } 1-\delta$$

◇ Replacing $\rho_N(\delta) \sim \sqrt{\gamma_N}$

$$\mathcal{R}(N) = \tilde{\mathcal{O}}\left(\gamma_N\sqrt{N}\right)$$

◇ Trivial ($\mathcal{O}(N)$): $\textit{Matérn}$ ($\nu \leq d/2$), $\textit{Laplace}$, $\textit{NTK}$

Discussion

◇ Could the square root of the effective dimension of the kernel in the regret bound be traded off for a square root of the input dimension?

$$\sqrt{\gamma_N} \quad \rightarrow \quad \sqrt{d \log(N)}$$

◇ The resulting regret bound

$$\mathcal{R}(N) = \tilde{\mathcal{O}}(\sqrt{d N \gamma_N})$$

◇ *SupKernelUCB* [Valko et al., 2013]: $\mathcal{R}(N) = \tilde{\mathcal{O}}(\sqrt{N \gamma_N})$, when $|\mathcal{X}| < \infty$

◇ Discretization argument: contributing only $\mathcal{O}(\sqrt{d \log(N)})$ factor

Special Case of Linear Models

Theorem [Abbasi-Yadkori et al., 2011] When $f = \mathbf{w}^\top x$ and $\|\mathbf{w}\|_{l^2} \leq B$, in the online setting, with probability $1-\delta$, for all $x \in \mathcal{X}$

$$|f(x) - \mu_n(x)| \leq \rho_n(\delta)\sigma_n(x)$$

- $\rho_n(\delta) = B + \frac{R}{\lambda}\sqrt{d \log(\frac{1+n\bar{x}^2/\lambda^2}{\delta})}$ and $\bar{x} = \max_{x \in \mathcal{X}} \|x\|_{\ell^2}$

Special Case of Linear Models

Theorem [Abbasi-Yadkori et al., 2011] When $f = \mathbf{w}^\top x$ and $\|\mathbf{w}\|_{l^2} \leq B$, in the online setting, with probability $1-\delta$, for all $x \in \mathcal{X}$

$$|f(x) - \mu_n(x)| \leq \rho_n(\delta)\sigma_n(x)$$

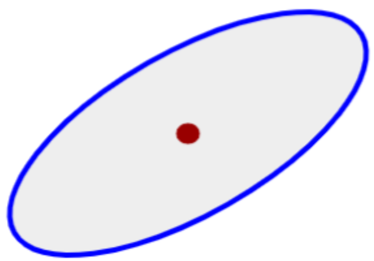
- $\rho_n(\delta) = B + \frac{R}{\lambda}\sqrt{d \log(\frac{1+n\bar{x}^2/\lambda^2}{\delta})}$ and $\bar{x} = \max_{x\in\mathcal{X}} \|x\|_{\ell^2}$

◇ Self-normalized bound for vector valued martingales $S_n = \sum_{i=1}^n \epsilon_i x_i$

◇ Confidence ellipsoid for $\mathbf{w}$:

$$\|\mathbf{w} - \hat{\mathbf{w}}_n\|_{V_n} \leq \lambda\rho_n(\delta) \text{ w.p. } 1-\delta$$

- $V_n = \lambda^2 \mathbf{I}_d + \sum_{i=1}^n x_i x_i^\top$



Online vs Offline Setting

◇ Online Setting: x_{n+1} is determined after $\{x_i, y_i\}_{i=1}^n$ are revealed

◇ Offline Setting: x_n is independent of all ϵ_i

Offline Confidence Intervals

◇ For a fixed $x \in \mathcal{X}$

$$|f(x) - \mu_n(x)| \leq \rho_0(\delta)\sigma_n(x), \text{ w.p. } 1-\delta$$

- $\rho_0(\delta) = B + \frac{R}{\lambda}\sqrt{2\log(\frac{2}{\delta})}$.

◇ When f is *Lipschitz* (or *Hölder*) continuous, uniformly in x,

$$|f(x) - \mu_n(x)| = \mathcal{O}\Big(\big(B + \frac{R}{\lambda}\sqrt{d\log(n) + \log(\frac{1}{\delta})}\big)\sigma_n(x)\Big) \text{ w.p. } 1-\delta$$

References

Y. Abbasi-Yadkori. Online learning for linearly parametrized control problems. 2013.

Y. Abbasi-Yadkori, D. Pl, and C. Szepesvri. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 2312–2320, 2011.

S. R. Chowdhury and A. Gopalan. On kernelized multi-armed bandits. In *Proceedings of International Conference on Machine Learning*, pages 844–853, 2017.

V. Nguyen, S. Gupta, S. Rana, C. Li, and S. Venkatesh. Regret for expected improvement over the best-observed value and stopping condition. In *Asian Conference on Machine Learning*, pages 279–294, 2017.

N. Srinivas, A. Krause, S. Kakade, and M. Seeger. Gaussian process optimization in the bandit setting: no regret and experimental design. In *International Conference on Machine Learning*, pages 1015–1022, 2010.

M. Valko, N. Korda, R. Munos, I. Flaounas, and N. Cristianini. Finite-time analysis of kernelised contextual bandits. In *Conference on Uncertainty in Artificial Intelligence*, pages 654–663, 2013.

Z. Yang, C. Jin, Z. Wang, M. Wang, and M. Jordan. Provably efficient reinforcement learning with kernel and neural function approximations. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

W. Zhang, D. Zhou, L. Li, and Q. Gu. Neural Thompson sampling. In *International Conference on Learning Representations*, 2021.

D. Zhou, L. Li, and Q. Gu. Neural contextual bandits with UCB-based exploration. In *International Conference on Machine Learning*, pages 11492–11502, 2020.