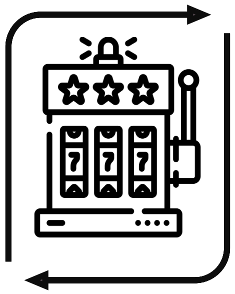
Neural Contextual Bandits without Regret

Parnian Kassraie, Andreas Krause

ETH zürich

Learning & Adaptive Systems

The Bandit Problem Setting



context $\boldsymbol{z}_t \in \mathcal{D}$

action $\boldsymbol{a} \in \mathcal{A}$

reward $y_t = f(\boldsymbol{x}_t) + \varepsilon_t$

$\boldsymbol{x}_t = (\boldsymbol{z}_t, \boldsymbol{a}) \in \mathcal{X}$ $\quad \mathcal{X} \subset \mathbb{S}^{d-1}$

$\varepsilon_t : \sigma^2$ sub-Gaussian

$f \in \mathcal{H}_k$

$\|f\|_k \leq B$

$\boldsymbol{k}$: the (Convolutional) Neural Tangent Kernel (NTK)

$\boldsymbol{\mathcal{H}_k}$: its Reproducing Kernel Hilbert Space (RKHS)

regret $R_T = \sum_{t=1}^{T} f(\boldsymbol{x}_t^*) - f(\boldsymbol{x}_t)$

$$\boldsymbol{x}_t^* = \arg\max_{\substack{\boldsymbol{x}=(\boldsymbol{z}_t, \boldsymbol{a}) \\ \boldsymbol{a} \in \mathcal{A}}} f(\boldsymbol{x})$$

goal $R_T/T \to 0$ as $T \to \infty$

Our algorithms: NN-UCB & CNN-UCB

How to pick the next action to control the regret?

Use any NN or 2-Layer CNN

train the network to estimate the reward

$$\min_{\boldsymbol{\theta}} \sum_{i<t} (f(\boldsymbol{x}_i; \boldsymbol{\theta}) - y_i)^2 + \sigma^2 \|\boldsymbol{\theta} - \boldsymbol{\theta}^0\|_2^2$$

$$\hat{\mu}_{t-1}(\boldsymbol{x}) = f(\boldsymbol{x}; \boldsymbol{\theta}^{(J)})$$ at step J of GD

use gradient of that network to estimate the variance of the reward

$$\boldsymbol{g}(\boldsymbol{x}) = \nabla_\theta f(\boldsymbol{x}; \boldsymbol{\theta}^{(0)})$$

$$\hat{\sigma}^2_{t-1} = \boldsymbol{g}^T(\boldsymbol{x}) \big[\sigma^2 \boldsymbol{I} + \sum_{i<t} \boldsymbol{g}^T(\boldsymbol{x}_i) \boldsymbol{g}(\boldsymbol{x}_i)\big]^{-1} \boldsymbol{g}(\boldsymbol{x})$$

$$\boldsymbol{x}_t = \arg\max \hat{\mu}_{t-1}(\boldsymbol{x}) + \sqrt{\beta_t} \hat{\sigma}_{t-1}(\boldsymbol{x})$$

Main Result: Sup(C)NN-UCB finds the optima in polynomial time.

Theorem (Informal)

Assume $f \in \mathcal{H}_{k_{NN}}$ (or $\mathcal{H}_{k_{CNN}}$). If

the learning rate is small enough

and the network is wide enough (or has many channels),

then under appropriate choice of β_t, (C)NN-UCB satisfies, (Sup Variant of)

$$R_T/T \to 0 \text{ as } T \to \infty$$

with high probability.

NN-UCB $\quad R(T)/T = \tilde{\mathcal{O}}\left(C_{\mathrm{NN}}(d,L)\, T^{\frac{-1}{2d}}\right)$

CNN-UCB $\quad R(T)/T = \tilde{\mathcal{O}}\left(\frac{C_{\mathrm{NN}}(d,L)}{d^{\frac{d-1}{2d}}}\, T^{\frac{-1}{2d}}\right)$

d: dimension of the input domain

Comparison to prior works

[Zhou et al. ICML '20]

[Zhang et al. ICLR '21]

[Yang et al. arXiv '20]

$$R_T \le \tilde{\mathcal{O}}\Big(I(\boldsymbol{y}_T; \boldsymbol{f}_T)\sqrt{T}\Big)$$

$I(\boldsymbol{y}_T; \boldsymbol{f}_T) = \tilde{\mathcal{O}}\left(T^{\frac{d-1}{d}}\right)$ [Thm 3.1]

$R_T/T = \tilde{\mathcal{O}}(T^{\frac{d-2}{2d}})$

Key Ingredient: Maximum Information Gain Bound

The information gain

$$I(\boldsymbol{y}_T; \boldsymbol{f}_T) = H(\boldsymbol{y}_T) - H(\boldsymbol{y}_T|\boldsymbol{f}_T) = \frac{1}{2}\log\det(\boldsymbol{I} + \sigma^{-2}\boldsymbol{K}_T)$$

$$\boldsymbol{y}_T = (y_1, \cdots, y_T)$$

$$\boldsymbol{f}_T = \big(f(\boldsymbol{x}_1), \cdots, f(\boldsymbol{x}_T)\big)$$

Its maximum

$$\gamma_T = \max_{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_T} I(\boldsymbol{y}_T; \boldsymbol{f}_T)$$

depends only on the domain, noise and kernel function

Theorem (Informal)

The maximum information gain associated with the Tangent Kernel of a L-layer NN (or a 2-layer CNN) is bounded by

$$\gamma_{T,NN} = \tilde{O}\left(C_{NN}(d, L) T^{\frac{d-1}{d}}\right)$$

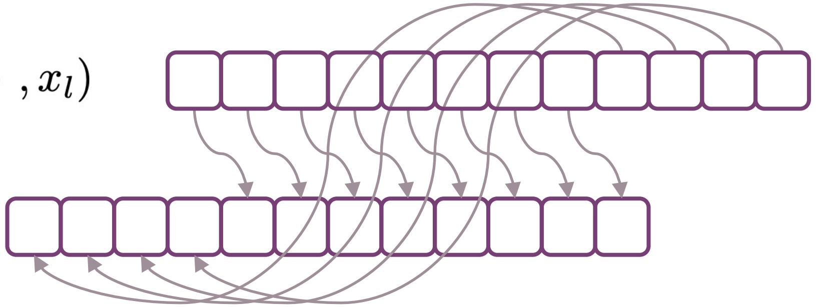
$$\gamma_{T,CNN} = \tilde{O}\left(C_{NN}(d, 2) \left(\frac{T}{d}\right)^{\frac{d-1}{d}}\right)$$

Key Ingredient II: Invariance Trick

Observation

$$\boldsymbol{w} * \boldsymbol{x} = \sum_{l=1}^{d} \langle \boldsymbol{w}, c_l \cdot \boldsymbol{x} \rangle$$

$$c_l \cdot \boldsymbol{x} = (x_{l+1}, x_{l+2}, \cdots, x_d, x_1, \cdots, x_l)$$



The 2-layer CNN is invariant to circular shifts

$$f_{\text{CNN}}(\boldsymbol{x}; \boldsymbol{W}, \boldsymbol{v}) = \frac{1}{d} \sum_{i=1}^{m} v_i \, \sigma_{\text{relu}}(\boldsymbol{w}_i * \boldsymbol{x}) = \frac{1}{d} \sum_{l=1}^{d} f_{\text{NN}}(c_l \cdot \boldsymbol{x}; \boldsymbol{W}, \boldsymbol{v})$$

And so is the corresponding CNTK

$$k_{\text{CNN}}(\boldsymbol{x}, \boldsymbol{x}') = \frac{1}{d} \sum_{l=1}^{a} k_{\text{NN}}(\boldsymbol{x}, c_l \cdot \boldsymbol{x}')$$

Key Ingredient II: Invariance Trick

On the d-1 dimensional sphere,

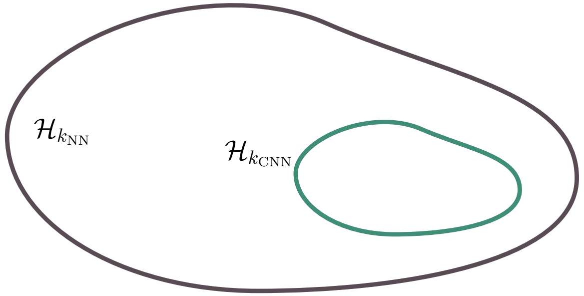
k_{NN} $\quad (\mu_k, \mathcal{F}_{d,k})$ $\quad$ spanned by degree-k spherical harmonics

(eigenvalue, eigenspace) pairs for $k \geq 0$

k_{CNN} $\quad (\mu_k, \bar{\mathcal{F}}_{d,k})$ $\quad$ spanned by circular shift invariant degree-k spherical harmonics

Lemma (Informal)

$$\frac{\dim\left(\bar{\mathcal{F}}_{d,k}\right)}{\dim\left(\mathcal{F}_{d,k}\right)} = \frac{1}{d}$$



→ **Improved rates for the CNN-UCB**

Thank you!