Combinatorial Gaussian Process Bandits with Probabilistically Triggered Arms

Ilker Demirel, Cem Tekin

Department of Electrical and Electronics Engineering
Bilkent University, Ankara, Turkey

This work was supported in part by the Scientific and Technological Research Council of Turkey (TUBITAK) under Grant 215E342 and in part by the BAGEP Award of the Science Academy.

CMAB-PTA

- We consider combinatorial multi-armed bandit (CMAB) with probabilistically triggered arms (PTAs) under the semi bandit feedback.
- e.g., cascading bandit and influence maximization bandits
- Classical Thompson sampling (TS) and upper confidence bound (UCB) based algorithms do not take correlations between base arms into account.
- We use Gaussian processes (GPs) to model base arm outcomes and propose Combinatorial GP-UCB (ComGP-UCB).
- ComGP-UCB enjoys sublinear regret and significantly outperforms classical TS and UCB methods when base arms are correlated.

Problem Formulation

Ilker Demirel,
Cem Tekin

- The learner chooses a subset of m base arms denoted by $S(t) \in \mathcal{I}$ at each round t where $\mathcal{I} \subseteq 2^{[m]}$ denotes the set of feasible super arms.
- Each base arm $i \in \{1, \ldots, m\}$ has a context denoted by $x_i \in \mathcal{X}$.
- We assume there exists $f : \mathcal{X} \to \mathbb{R}$ where f is sampled from a GP and $f(x_i) = \mu_i$.

ComGP-UCB

Ilker Demirel,
Cem Tekin

- ComGP-UCB forms an estimate of the expected base-arm outcomes, $\bar{\boldsymbol{\mu}}_t$
- The oracle knows the problem structure and plays the optimal super arm S based on $\bar{\boldsymbol{\mu}}_t$
- Once the feedback is observed, ComGP-UCB updates the posterior distribution of GP –exploiting the relevance between different arms– before the next round.

ComGP-UCB

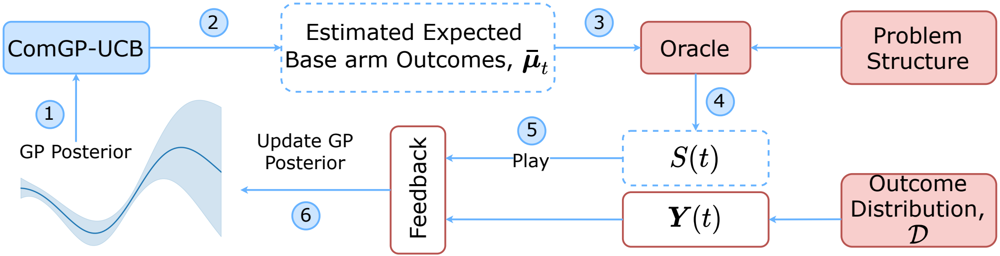

Figure: ComGP-UCB

Ilker Demirel,
Cem Tekin

Theorem

Given Lipschitz constant B, δ, $\rho \in (0,1)$, $\alpha \in [mT]$, under Assumptions 1, 2, and 3, and when $\sqrt{T} > \frac{m + \frac{2m}{\rho^2}\mathbb{E}_{\mu}[\frac{1}{p^*}]}{\delta}$, the cumulative regret of ComGP-UCB after round T is upper bounded with at least $1 - 2\delta$ probability as follows,

$$\mathbb{P}\{Reg_{\mu}(T) \leq 4mB\sqrt{\frac{T\beta_{mT}\sigma^2}{(1-\rho)p^*}} + 2m\alpha B\sqrt{\beta_{mT}}\} \geq 1 - 2\delta.$$

Ilker Demirel,
Cem Tekin

Theoretical Bounds

Theorem

Given $\delta \in (0,1)$, $C := \frac{8\tilde{B}^2}{\log(1+\sigma^{-2})}$, *where* $\tilde{B}$ *is the TPM Lipschitz constant, and under Assumptions 1, 2, and 4, the cumulative regret of ComGP-UCB after round* T *is upper bounded with at least* $1-\delta$ *probability as follows,*

$$\mathbb{P}\{Reg_{\mu}(T) \leq \sqrt{CmT\beta_{mT}\gamma_{T,\mu}^{PTA}}\} \geq 1-\delta.$$

Comparison with Related Work

Ilker Demirel, Cem Tekin

Table 1: Our Work in Comparison to Related Work

ALGO.	PUBL.	Regret Bound
CUCB	(Chen et al., 2013)	$O(\sum_i \log T/\Delta_i)$
CUCB[1]	(Chen et al., 2016)	$O(\sum_i \log T/(p_i\Delta_i))$
CUCB[1]	(Wang et al., 2017)	$O(\sum_i \log T/\Delta_i)^*$
CTS[2]	(Wang et al., 2018)	$O(\sum_i \log T/\Delta_i)$
CTS[1,2]	(Hüyük et al., 2019)	$O(\sum_i \log T/(p_i\Delta_i))$
ComGP-UCB[1,2]		$O\left(m\sqrt{\frac{T\log T}{p^*}}\right)$ $O\left(\sqrt{mT\log T\gamma_{T,\boldsymbol{\mu}}^{PTA}}\right)^*$

*With triggering probability modulated Lipschitz continuity
[1]PTA scenario considered
[2]Exact oracle used

Experiments

- Cascading bandit problem (item list recommendation)
- A synthetic linear problem

Cascading Bandit

Ilker Demirel,
Cem Tekin

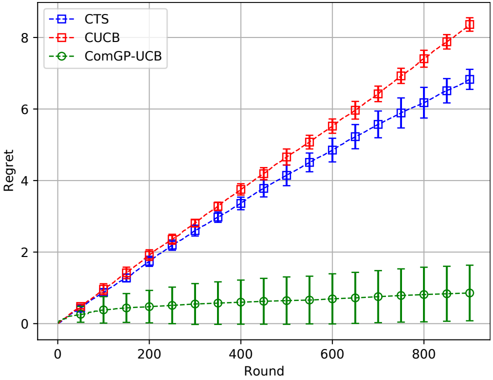

Figure: High Correlation Scenario

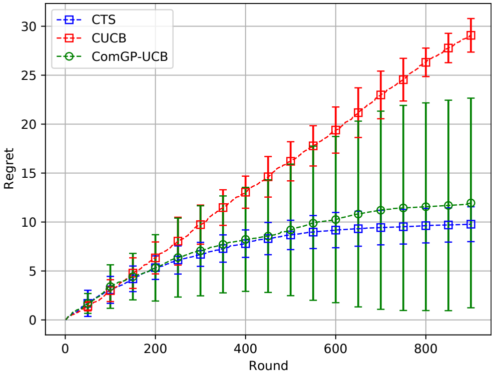

Figure: No Correlation Scenario

Synthetic Linear Problem

Ilker Demirel,
Cem Tekin