Paper Reading Assignment 3

HU Yang

Parallelizing Exploration-Exploitation Tradeoffs in Gaussian Process Bandit Optimization

SDSC8014 – Online learning

March 12, 2022

1 Introduction

This paper considers an upper confidence bound-based algorithm for the parallelizing exploration-exploitation problem in Gaussian Process Bandit Optimization. The core motivation of this paper is to *parallelize the numerical experiments* to collect more information of the same time frame. Still, there are two remaining challenges in academia:

- Choosing the batch of experiments are computational costly.
- The size of batches impact algorithmic performance.

In this paper, the authors stated multiple contributions:

- Gaussian Process Batch Upper Confidence Bound (GP-BUCB) is proposed, where Gaussian Process Upper Confidence Bound (GP-UCB) decision rule is adopted in each batch.
- An adaptive parallelism based Gaussian Process Adaptive Upper Confidence Bound (GP-AUCB) algorithm is proposed.
- The accumulative regret of sub-linear bounds is provided.
- By parallelizing the computing process, the asymptotic convergence rate is near-linear.
- If we evaluating the posterior variance by lazy variance calculations, GP-BUCB and GP-AUCB can be accelerated.
- The performance of GP-BUCB and GP-AUCB is proved on various scenarios

2 Preliminaries

First of all, the basic notations are defined:

- Given a series of decisions: x_1, x_2,..., x_T
- The decision set is denoted as D
- The payoff function, f, is unknown
- Incorporating observation noise

- Regret at time slot t is defined as the difference of the payoff function on offline optimal and x_t. If the average regret is sub-linear, it will be approximated to zero if T goes to infinity

There are two ways for parallelization: pessimistic and optimistic. The paper concentrate on pessimistic view, for the benchmark in optimistic view is ambiguous. Also, there are two settings of feedback mapping:

- Simple batch setting:

$$\text{fb}[t]_{SB} = \begin{cases} 0 & : t \in \{1, \ldots, B\} \\ B & : t \in \{B+1, \ldots, 2B\} \\ 2B & : t \in \{2B+1, \ldots, 3B\} \\ \vdots & \end{cases}$$

- Simple delay setting:

$$\text{fb}[t]_{SD} = \max(t - B, 0)$$

The way we model payoff function f is based on *Gaussian process*, where f is sampled from a *Gaussian process* prior probability distribution. The posterior distribution of f is:

$$f(\boldsymbol{x})|\boldsymbol{y}_{1:t-1} \sim \mathcal{N}(\mu_{t-1}(\boldsymbol{x}), \sigma_{t-1}^2(\boldsymbol{x}))$$

The mean and variance in different time slot are defined as:

$$\mu_{t-1}(\boldsymbol{x}) = K(\boldsymbol{x}, \boldsymbol{X}_{t-1})[K(\boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-1}) + \sigma_n^2 I]^{-1}\boldsymbol{y}_{1:t-1} \text{ and}$$
$$\sigma_{t-1}^2(\boldsymbol{x}) = k(\boldsymbol{x}, \boldsymbol{x}) - K(\boldsymbol{x}, \boldsymbol{X}_{t-1})[K(\boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-1}) + \sigma_n^2 I]^{-1}K(\boldsymbol{x}, \boldsymbol{X}_{t-1})^T$$

where $k(x_i, x_j)$ is the kernel function.

$$[K(\boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-1})]_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j),\ \forall \boldsymbol{x}_i, \boldsymbol{x}_j \in \boldsymbol{X}_{t-1}$$

The mutual information between the function f and the observation y is defined as:

$$I(f; \boldsymbol{y}_A) = H(\boldsymbol{y}_A) - H(\boldsymbol{y}_A | f) = \frac{1}{2}\log\left|\mathbf{I} + \sigma_n^{-2}K(A, A)\right|$$

where $K(A, A)$ is the covariance matrix of the values of f at the element of A, and H is the differential entropy of the probability distribution.

3 The GP-BUCB algorithm

The framework of GP-BUCB is presented below:

Algorithm 2 GP-BUCB

Input: Decision set D, GP prior μ_0, σ_0, kernel function $k(\cdot,\cdot)$, feedback mapping $\text{fb}[\cdot]$.
for $t = 1, 2, \ldots, T$ **do**
  Choose $\boldsymbol{x}_t = \text{argmax}_{\boldsymbol{x} \in D}[\mu_{\text{fb}[t]}(\boldsymbol{x}) + \beta_t^{1/2}\sigma_{t-1}(\boldsymbol{x})]$
  Compute $\sigma_t(\cdot)$ via Equation (2)
  if $\text{fb}[t] < \text{fb}[t+1]$ **then**
    Obtain $y_{t'} = f(\boldsymbol{x}_{t'}) + \varepsilon_{t'}$ for $t' \in \{\text{fb}[t]+1, \ldots, \text{fb}[t+1]\}$
    Perform Bayesian inference to obtain $\mu_{\text{fb}[t+1]}(\cdot)$ via Equation (1)
  end if
end for

As we see in the first step of the iteration in GP-BUCB, x_t is chosen with high mean and high variance:

$$\boldsymbol{x}_t = \underset{\boldsymbol{x}\in D}{\operatorname{argmax}} \left[\mu_{t-1}(\boldsymbol{x}) + \alpha_t^{1/2}\sigma_{t-1}(\boldsymbol{x})\right]$$

Thus, the average regret approaches 0, where α_T is the confidence interval growing poly-logarithmically in T, and γ_T grows sublinear in T:

$$R_T = O(\sqrt{T\alpha_T\gamma_T})$$

Well, GP-BUCB has an edge on encouraging diversity in exploration, while the predictive variance depends on the assumed observations. However, it still has some drawbacks. The algorithm will be over-confident if we know more of actual state. Further, it does not explore sufficiently because f changes smoothly, and it suffers from linear regret.
The regret bound of GP-BUCB is proved in a corollary, which is independent of the batch size, B:

Corollary 3 *Assume the GP-BUCB algorithm is employed with a constant B such that $t - fb[t] \leq B$ for all $t \geq 1$. Let $\delta \in (0,1)$, and let the requirements of one of the numbered cases of Theorem 2 be met. Choose $\beta_t = \exp(2C)\alpha_{fb[t]+1}$ (Cases 1 & 3) or $\beta_t = \exp(2C)\alpha_t$ (Case 2) and select actions $\boldsymbol{x}_t$ for all $t \geq 1$. Then*

$$\Pr\left\{R_T \leq \sqrt{C_1 T \exp(2\gamma_{B-1})\alpha_T\gamma_T} + 2, \quad \forall T \geq 1\right\} \geq 1-\delta,$$

where $C_1 = 8/\log(1+\sigma_n^{-2})$ and γ_{B-1} and γ_τ are as defined in Equation (4).

4 The GP-AUCB algorithm

GP-AUCB algorithm adaptively chooses fb[t] to control the batch size. The current batch B is dynamic. If the information is more than the pre-defined threshold constant C, the algorithm will do the following changes:

Algorithm 4 GP-AUCB

Input: Decision set D, GP prior μ_0, σ_0, kernel $k(\cdot,\cdot)$, information gain threshold C.
Set $\text{fb}[t'] = 0$, $\forall t' \geq 1$, $G = 0$.
for $t = 1, 2, \ldots, T$ **do**
  if $G > C$ **then**
    Obtain $y_{t'} = f(\boldsymbol{x}_{t'}) + \varepsilon_{t'}$ for $t' \in \{\text{fb}[t-1], \ldots, t-1\}$
    Perform Bayesian inference to obtain $\mu_{t-1}(\cdot)$ via Equation (1)
    Set $G = 0$
    Set $\text{fb}[t'] = t-1$, $\forall t' \geq t$
  end if
  Choose $\boldsymbol{x}_t = \text{argmax}_{\boldsymbol{x}\in D}[\mu_{\text{fb}[t]}(\boldsymbol{x}) + \beta_t^{1/2}\sigma_{t-1}(\boldsymbol{x})]$
  Set $G = G + \frac{1}{2}\log(1 + \sigma_n^{-2}\sigma_{t-1}^2(\boldsymbol{x}_t))$
  Compute $\sigma_t(\cdot)$ via Equation (2)
end for

The regret of GP-AUCB is proved in Corollary 6:

Corollary 6 *Let the GP-AUCB algorithm be employed with a specified constant* $\delta \in (0,1)$ *and a specified constant* $C > 0$*, for which the resulting feedback mapping* $fb : \mathbb{N} \to \mathbb{N}$ *guarantees* $I(f; \boldsymbol{y}_{fb[t]+1:t-1} \mid \boldsymbol{y}_{1:fb[t]}) \leq C, \forall t \geq 1$*. If the conditions of one case of Theorem 2 are met, and* $\beta_t = \exp(2C)\alpha_{fb[t]+1}$ *(Case 1 & 3) or* $\beta_t = \exp(2C)\alpha_t$ *(Case 2), then*

$$\Pr\left\{R_T \leq \sqrt{C_1 T \exp(2C)\alpha_T \gamma_T} + 2, \forall T \geq 1\right\} \geq 1 - \delta$$

where $C_1 = 8/\log(1 + \sigma_n^{-2})$.

5 Lazy variance calculation

Lazy variance calculation is used to accelerate UCB-based algorithms without loss of performance. The basic idea of Lazy variance calculation is maintain one upper bound variance instead of calculating the variance every round. Then, using this upper bound to select the optimal action.

$$\boldsymbol{x}_t = \underset{\boldsymbol{x} \in D}{\operatorname{argmax}} \left[\mu_{fb[t]}(\boldsymbol{x}) + \beta_t^{1/2} \widetilde{\sigma}_{t-1}(\boldsymbol{x})\right]$$