Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design

Rudenko Varvara

MIPT
MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY

April, 2021

Summary

1. Abstract
2. Summary
3. Problem Statement and Background
4. Experiments
5. Conclusions

Abstract

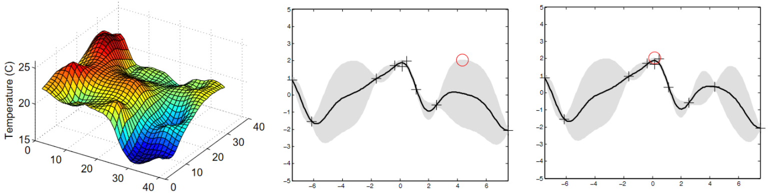
- Many applications require optimizing an unknown, noisy function that is expensive to evaluate. We formalize this task as a multi-armed bandit problem, where the payoff function is either sampled from a Gaussian process (GP) or has low RKHS norm.
- We resolve the important open problem of deriving regret bounds for this setting, which imply novel convergence rates for GP optimization. We analyze GP-UCB, an intuitive upper-confidence based algorithm, and bound its cumulative regret in terms of maximal information gain, establishing a novel connection between GP optimization and experimental design.

Summary

- We analyze GP-UCB, an intuitive algorithm for GP optimization, when the function is either sampled from a known GP, or has low RKHS norm.
- We bound the cumulative regret for GP-UCB interms of the information gain due to sampling, establishing a novel connection between experimental design and GP optimization.
- By bounding the information gain for popular classes of kernels, we establish sublinear regret bounds for GP optimization for the first time.
- We evaluate GP-UCB on sensor network data, demonstrating that it compares favorably to existing algorithms for GP optimization.

Problem Statement and Background

- For example, we might want to find locations of highest temperature in a building by sequentially activating sensors in a spatial network and regressing on their measurements. D consists of all sensor locations, f(x) is the temperature at x, and sensor accuracy is quantified by the noise variance. Each activation draws battery power, so we want to sample from as few sensors as possible.



(a) *Temperature data* (b) *Iteration t* (c) *Iteration $t+1$*

Figure: (a) Example of temperature data collected by a network of 46 sensors at Intel Research Berkeley.

$$x_t = argmax_{x \in D} \sigma_{t-1}(x) \tag{1}$$

$$x_t = argmax_{x \in D} \mu_{t-1}(x) \tag{2}$$

$$x_t = argmax_{x \in D} \mu_{t-1}(x) + {\beta_t}^{\frac{1}{2}} \sigma_{t-1}(x) \tag{3}$$

Algorithm 1 The GP-UCB algorithm.

Input: Input space D; GP Prior $\mu_0 = 0,\ \sigma_0,\ k$

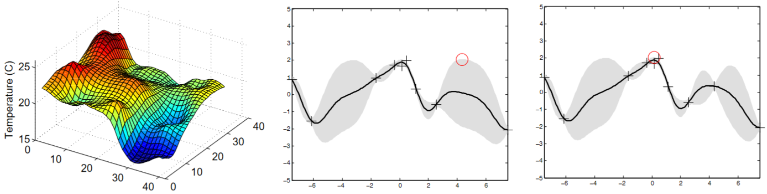
for $t = 1, 2, \ldots$ **do**

  Choose $\boldsymbol{x}_t = \underset{\boldsymbol{x} \in D}{\operatorname{argmax}}\, \mu_{t-1}(\boldsymbol{x}) + \sqrt{\beta_t}\sigma_{t-1}(\boldsymbol{x})$

  Sample $y_t = f(\boldsymbol{x}_t) + \epsilon_t$

  Perform Bayesian update to obtain μ_t and σ_t

end for

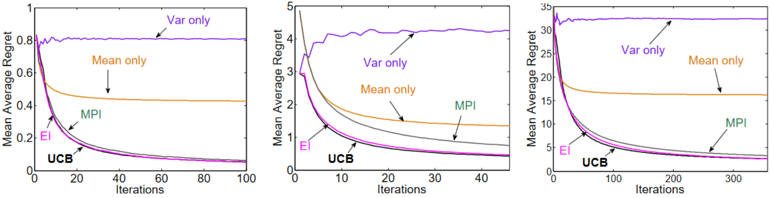


(a) *Temperature data* (b) *Iteration t* (c) *Iteration $t+1$*

Figure: (b,c) Two iterations of the GP-UCB algorithm. It samples points that are either uncertain (b) or have high posterior mean (c).

Experiments

- We compare GP-UCB with heuristics such as the Expected Improvement (EI) and Most Probable Improvement (MPI), and with naive methods which choose points of maximum mean or variance only, both on synthetic and real sensor network data.



(a) *Squared exponential* (b) *Temperature data* (c) *Traffic data*

Figure: Comparison of performance: GP-UCB and various heuristics on synthetic (a), and sensor network data (b, c).

Conclusions

- We analyze GP-UCB, an intuitive, Bayesian upper confidence bound based sampling rule.Our regret bounds crucially depend on the information gain due to sampling, establishing a novel connection between bandit optimization and experimental design. Our results provide an interesting step towards understanding exploration–exploitation trade offs with complex utility functions.

Thank you!