

# Active Learning is Planning: Nonmyopic $\epsilon$ -Bayes-Optimal Active Learning of Gaussian Processes

Trong Nghia Hoang<sup>1</sup>, Kian Hsiang Low<sup>1</sup>, Patrick Jaillet<sup>2</sup>, and Mohan Kankanhalli<sup>1</sup>

<sup>1</sup> National University of Singapore {nghiaht, lowkh, mohan}@comp.nus.edu.sg

<sup>2</sup> Massachusetts Institute of Technology jaillet@mit.edu

**Abstract.** A fundamental issue in active learning of Gaussian processes is that of the exploration-exploitation trade-off. This paper presents a novel nonmyopic  $\epsilon$ -Bayes-optimal active learning ( $\epsilon$ -BAL) approach [4] that jointly optimizes the trade-off. In contrast, existing works have primarily developed greedy algorithms or performed exploration and exploitation separately. To perform active learning in real time, we then propose an anytime algorithm [4] based on  $\epsilon$ -BAL with performance guarantee and empirically demonstrate using a real-world dataset that, with limited budget, it outperforms the state-of-the-art algorithms.

## 1 Introduction

Active learning/sensing has become an increasingly important focal theme in environmental sensing and monitoring applications (e.g., precision agriculture [7], monitoring of ocean and freshwater phenomena). Its objective is to derive an optimal sequential policy that plans the most informative locations to be observed for minimizing the predictive uncertainty of the unobserved areas of a spatially varying environmental phenomenon given a sampling budget (e.g., number of deployed sensors, energy consumption). To achieve this, many existing active sensing algorithms [1, 2, 3, 6, 7, 8] have modeled the phenomenon as a *Gaussian process* (GP), which allows its spatial correlation structure to be formally characterized and its predictive uncertainty to be formally quantified (e.g., based on entropy, or mutual information criterion). However, they have assumed the spatial correlation structure (specifically, the parameters defining it) to be known, which is often violated in real-world applications. The predictive performance of the GP model in fact depends on how informative the gathered observations are for both parameter estimation and spatial prediction given the true parameters.

Interestingly, as revealed in [9], policies that are efficient for parameter estimation are not necessarily efficient for spatial prediction with respect to the true model parameters. Thus, active learning/sensing involves a potential trade-off between sampling the most informative locations for spatial prediction given the current, possibly incomplete knowledge of the parameters (i.e., exploitation) vs. observing locations that gain more information about the parameters (i.e., exploration). To address this trade-off, one principled approach is to frame active sensing as a sequential decision problem that jointly optimizes the above exploration-exploitation trade-off while maintaining a Bayesian belief over the model parameters. Solving this problem then results in an induced policy that is guaranteed to be optimal in the expected active sensing performance [4]. Unfortunately, such a nonmyopic *Bayes-optimal active learning* (BAL) policy cannot be derived exactly due to an uncountable set of candidate observations and unknown model parameters. As a result, existing works advocate using greedy policies [10] or

performing exploration and exploitation separately [5] to sidestep the difficulty of solving for the exact BAL policy. But, these algorithms are sub-optimal in the presence of budget constraints due to their imbalance between exploration and exploitation [4].

This paper presents a novel nonmyopic active learning algorithm [4] that can still preserve and exploit the principled Bayesian sequential decision problem framework for jointly optimizing the exploration-exploitation trade-off (Section 2.2) and consequently does not incur the limitations of existing works. In particular, although the exact BAL policy cannot be derived, we show that it is in fact possible to solve for a nonmyopic  $\epsilon$ -Bayes-optimal active learning ( $\epsilon$ -BAL) policy (Section 2.3) given an arbitrary loss bound  $\epsilon$ . To meet real-time requirement in time-critical applications, we then propose an asymptotically  $\epsilon$ -optimal anytime algorithm based on  $\epsilon$ -BAL with performance guarantee (Section 2.4). We empirically demonstrate using a real-world dataset that, with limited budget, our approach outperforms state-of-the-art algorithms (Section 3).

## 2 Nonmyopic $\epsilon$ -Bayes-Optimal Active Learning

### 2.1 Modeling Spatial Phenomena with Gaussian Processes

Let  $\mathcal{X}$  denote a set of sampling locations representing the domain of the phenomenon such that each location  $x \in \mathcal{X}$  is associated with a realized (random) measurement  $z_x$  ( $Z_x$ ) if  $x$  is observed (unobserved). Let  $Z_{\mathcal{X}} \triangleq \{Z_x\}_{x \in \mathcal{X}}$  denote a GP [4]. The GP is fully specified by its *prior* mean  $\mu_x \triangleq \mathbb{E}[Z_x]$  and covariance  $\sigma_{xx'}|\lambda \triangleq \text{cov}[Z_x, Z_{x'}|\lambda]$  for all locations  $x, x' \in \mathcal{X}$ ; its model parameters are denoted by  $\lambda$ . When  $\lambda$  is known and a set  $z_{\mathcal{D}}$  of realized measurements is observed for  $\mathcal{D} \subset \mathcal{X}$ , the GP prediction for any unobserved location  $x \in \mathcal{X} \setminus \mathcal{D}$  is given by  $p(z_x|z_{\mathcal{D}}, \lambda) = \mathcal{N}(\mu_x|\mathcal{D}, \lambda, \sigma_{xx}|\mathcal{D}, \lambda)$  [4]. However, since  $\lambda$  is not known, a probabilistic belief  $b_{\mathcal{D}}(\lambda) \triangleq p(\lambda|z_{\mathcal{D}})$  is maintained over all possible  $\lambda$  and updated using Bayes' rule to the posterior belief  $b_{\mathcal{D} \cup \{x\}}(\lambda) \propto p(z_x|z_{\mathcal{D}}, \lambda) b_{\mathcal{D}}(\lambda)$  given a new measurement  $z_x$ . Then, using belief  $b_{\mathcal{D}}$ , the predictive distribution is obtained by marginalizing out  $\lambda$ :  $p(z_x|z_{\mathcal{D}}) = \sum_{\lambda \in \Lambda} p(z_x|z_{\mathcal{D}}, \lambda) b_{\mathcal{D}}(\lambda)$ .

### 2.2 Problem Formulation

To cast active sensing as a Bayesian sequential decision problem, we define a sequential active sensing policy  $\pi \triangleq \{\pi_n\}_{n=1}^N$  that is structured to sequentially decide the next location  $\pi_n(z_{\mathcal{D}}) \in \mathcal{X} \setminus \mathcal{D}$  to be observed at each stage  $n$  based on the current observations  $z_{\mathcal{D}}$  over a finite planning horizon of  $N$  stages (i.e., sampling budget). To measure the predictive uncertainty over unobserved areas of the phenomenon, we use the entropy criterion and define the value under a policy  $\pi$  to be the joint entropy of its selected observations when starting with some prior observations  $z_{\mathcal{D}_0}$  and following  $\pi$  thereafter [4]. The work of [7] has established that minimizing the posterior joint entropy (i.e., predictive uncertainty) remaining in unobserved locations of the phenomenon is equivalent to maximizing the joint entropy of  $\pi$ . Thus, solving the active sensing problem entails choosing a sequential BAL policy  $\pi_n^*(z_{\mathcal{D}}) = \arg \max_{x \in \mathcal{X} \setminus \mathcal{D}} Q_n^*(z_{\mathcal{D}}, x)$  induced from the following  $N$ -stage Bellman equations, as formally derived in [4]:

$$V_n^*(z_{\mathcal{D}}) \triangleq \max_{x \in \mathcal{X} \setminus \mathcal{D}} Q_n^*(z_{\mathcal{D}}, x) \quad (1)$$

$$Q_n^*(z_{\mathcal{D}}, x) \triangleq \mathbb{E}[-\log p(Z_x|z_{\mathcal{D}})] + \mathbb{E}[V_{n+1}^*(z_{\mathcal{D}} \cup \{Z_x\})|z_{\mathcal{D}}]$$

for stage  $n = 1, \dots, N$  where  $p(z_x|z_{\mathcal{D}})$  is defined in Section 2.1 and the second expectation term is omitted from right-hand side expression of  $Q_N^*$  at stage  $N$ . Unfortunately,

since the BAL policy  $\pi^*$  cannot be derived exactly, we instead consider solving for an  $\epsilon$ -BAL policy  $\pi^\epsilon$  whose joint entropy approximates that of  $\pi^*$  within  $\epsilon > 0$ .

### 2.3 $\epsilon$ -BAL Policy

The key idea of our proposed nonmyopic  $\epsilon$ -BAL policy  $\pi^\epsilon$  is to approximate the expectation terms in (1) at every stage using truncated sampling. Specifically, given realized measurements  $z_{\mathcal{D}}$ , a finite set of  $\tau$ -truncated, i.i.d. observations  $\{z_x^i\}_{i=1}^S$  [4] is generated and exploited for approximating  $V_n^*$  (1) through the following Bellman equations:

$$\begin{aligned} V_n^\epsilon(z_{\mathcal{D}}) &\triangleq \max_{x \in \mathcal{X} \setminus \mathcal{D}} Q_n^\epsilon(z_{\mathcal{D}}, x) \\ Q_n^\epsilon(z_{\mathcal{D}}, x) &\triangleq \frac{1}{S} \sum_{i=1}^S -\log p(z_x^i | z_{\mathcal{D}}) + V_{n+1}^\epsilon(z_{\mathcal{D}} \cup \{z_x^i\}) \end{aligned} \quad (2)$$

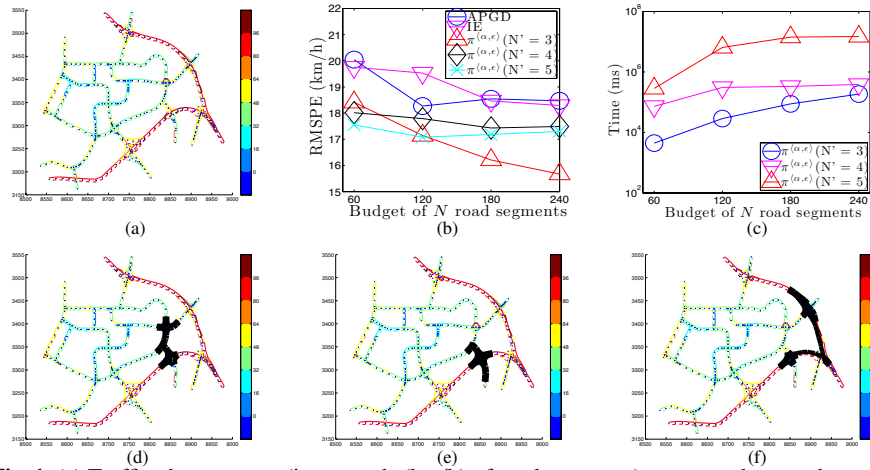
for stage  $n = 1, \dots, N$ . The use of truncation is motivated by a technical necessity for theoretically guaranteeing the *expected* active sensing performance (specifically,  $\epsilon$ -Bayes-optimality) of  $\pi^\epsilon$  relative to that of  $\pi^*$  [4].

### 2.4 Anytime $\epsilon$ -BAL ( $\langle \alpha, \epsilon \rangle$ -BAL) Algorithm

Although  $\pi^\epsilon$  can be derived exactly, the cost of deriving it is exponential in the length  $N$  of planning horizon since it has to compute the values  $V_n^\epsilon(z_{\mathcal{D}})$  (2) for all  $(S|\mathcal{X}|)^N$  possible states  $(n, z_{\mathcal{D}})$ . To ease this computational burden, we propose an anytime algorithm based on  $\epsilon$ -BAL that can produce a good policy fast and improve its approximation quality over time. The key intuition behind our *anytime  $\epsilon$ -BAL algorithm* ( $\langle \alpha, \epsilon \rangle$ -BAL) is to focus the simulation of greedy exploration paths through the most uncertain regions of the state space (i.e., in terms of the values  $V_n^\epsilon(z_{\mathcal{D}})$ ) instead of evaluating the entire state space like  $\pi^\epsilon$ . Interested readers are referred to [4] for more details.

## 3 Experiments and Discussion

This section evaluates the active sensing performance and time efficiency of our  $\langle \alpha, \epsilon \rangle$ -BAL policy  $\pi^{\langle \alpha, \epsilon \rangle}$  empirically under using a real-world dataset of a large-scale traffic phenomenon (i.e., speeds of road segments) over an urban road network; refer to [4] for additional experimental results on a simulated spatial phenomenon. Fig. 1a shows the urban road network  $\mathcal{X}$  comprising 775 road segments in Tampines area, Singapore during lunch hours on June 20, 2011. Each road segment  $x \in \mathcal{X}$  is specified by a 4-dimensional vector of features: length, number of lanes, speed limit, and direction. More details of our experimental setup can be found in [4]. The performance of our  $\langle \alpha, \epsilon \rangle$ -BAL policies with planning horizon length  $N' = 3, 4, 5$  are compared to that of APGD and IE policies [5] by running each of them on a mobile robotic probe to direct its active sensing along a path of adjacent road segments according to the road network topology. Fig. 1 shows results of the tested policies averaged over 5 independent runs: It can be observed from Fig. 1b that our  $\langle \alpha, \epsilon \rangle$ -BAL policies outperform APGD and IE policies due to their nonmyopic exploration behavior. Fig. 1c shows that  $\langle \alpha, \epsilon \rangle$ -BAL incurs  $< 4.5$  hours given a budget of  $N = 240$  road segments, which can be afforded by modern computing power. To illustrate the behavior of each policy, Figs. 1d-f show, respectively, the road segments observed (shaded in black) by the mobile probe running APGD, IE, and  $\langle \alpha, \epsilon \rangle$ -BAL policies with  $N' = 5$  given a budget of  $N = 60$ . Interestingly, Figs. 1d-e show that both APGD and IE cause the probe to move away from the slip roads and highways to low-speed segments whose measurements vary



**Fig. 1.** (a) Traffic phenomenon (i.e., speeds (km/h) of road segments) over an urban road network, graphs of (b) root mean squared prediction error of APGD, IE, and  $\langle \alpha, \epsilon \rangle$ -BAL policies with horizon length  $N' = 3, 4, 5$  and (c) total online processing cost of  $\langle \alpha, \epsilon \rangle$ -BAL policies with  $N' = 3, 4, 5$  vs. budget of  $N$  segments, and (d-f) road segments observed (shaded in black) by respective APGD, IE, and  $\langle \alpha, \epsilon \rangle$ -BAL policies ( $N' = 5$ ) with  $N = 60$ .

much more smoothly; this is expected due to their myopic exploration behavior. In contrast,  $\langle \alpha, \epsilon \rangle$ -BAL nonmyopically plans the probe's path and direct it to observe the more informative slip roads and highways with highly varying traffic measurements (Fig. 1f) to achieve better performance.

**Acknowledgments.** This work was supported by Singapore National Research Foundation in part under its International Research Center @ Singapore Funding Initiative and administered by the Interactive Digital Media Programme Office and in part through the Singapore-MIT Alliance for Research & Technology Subaward Agreement No. 52.

## Bibliography

- [1] Cao, N., Low, K.H., Dolan, J.M.: Multi-robot informative path planning for active sensing of environmental phenomena: A tale of two algorithms. In: Proc. AAMAS. pp. 7–14 (2013)
- [2] Chen, J., Low, K.H., Tan, C.K.Y.: Gaussian process-based decentralized data fusion and active sensing for mobility-on-demand system. In: Proc. RSS (2013)
- [3] Chen, J., Low, K.H., Tan, C.K.Y., Oran, A., Jaillet, P., Dolan, J.M., Sukhatme, G.S.: Decentralized data fusion and active sensing with mobile sensors for modeling and predicting spatiotemporal traffic phenomena. In: Proc. UAI. pp. 163–173 (2012)
- [4] Hoang, T.N., Low, K.H., Jaillet, P., Kankanhalli, M.: Nonmyopic  $\epsilon$ -Bayes-Optimal Active Learning of Gaussian Processes. In: Proc. ICML. pp. 739–747 (2014)
- [5] Krause, A., Guestrin, C.: Nonmyopic active learning of Gaussian processes: An exploration-exploitation approach. In: Proc. ICML. pp. 449–456 (2007)
- [6] Low, K.H., Dolan, J.M., Khosla, P.: Adaptive multi-robot wide-area exploration and mapping. In: Proc. AAMAS. pp. 23–30 (2008)
- [7] Low, K.H., Dolan, J.M., Khosla, P.: Information-theoretic approach to efficient adaptive path planning for mobile robotic environmental sensing. In: Proc. ICAPS (2009)
- [8] Low, K.H., Dolan, J.M., Khosla, P.: Active Markov information-theoretic path planning for robotic environmental sensing. In: Proc. AAMAS. pp. 753–760 (2011)
- [9] Martin, R.J.: Comparing and contrasting some environmental and experimental design problems. *Environmetrics* 12(3), 303–317 (2001)
- [10] Ouyang, R., Low, K.H., Chen, J., Jaillet, P.: Multi-robot active sensing of non-stationary Gaussian process-based environmental phenomena. In: Proc. AAMAS. pp. 573–580 (2014)