Self-Confidence in Autonomous Systems
Papers from the AAAI 2015 Fall Symposium

Uninformed-to-Informed Exploration in Unstructured Real-World Environments

Allan Axelrod and **Girish Chowdhary**
{allanma,girish.chowdhary}@okstate.edu
Oklahoma State University
Stillwater, Oklahoma 74075

Abstract

Conventionally, the process of learning the model (exploration) is initialized as either an uninformed or informed policy, where the latter leverages observations to guide future exploration. Informed exploration is ideal as it may allow a model to be learned in fewer samples. However, informed exploration cannot be implemented from the onset when a-priori knowledge on the sensing domain statistics are not available; such policies would only sample the first set of locations, repeatedly. Hence, we present a theoretically-derived bound for transitioning from uninformed exploration to informed exploration for unstructured real-world environments which may be partially-observable and time-varying. This bound is used in tandem with a sparsified Bayesian nonparametric Poisson Exposure Process, which is used to learn to predict the value of information in partially-observable and time-varying domains. The result is an uninformed-to-informed exploration policy which outperforms baseline algorithms in real-world data-sets.

Introduction

When can an agent become confident in its ability to predict where the most valuable information will be in some unstructured environment? This broad question is not theoretically treated in state-of-the-art information-theoretic learning approaches such as (Little and Sommer 2013; Russo and Van Roy 2014) which initialize as informed search strategies with an expert-defined duration of pre-training in structured environments. In contrast, we examine when a transition between uninformed and informed learning behaviors may successfully occur using a bound extended from the premise of the Bienaymé-Chebyshev bound (Heyde and Seneta 1972). The so-called Domain Exposure Bound allows the agent to confirm that it can engage in informed sampling with probabilistic guarantees on successfully optimizing the task of learning in unstructured environments. Herein, operation in an unstructured environment means that an agent is without prior information on the statistics or patterns in an environment with stationary, time-varying, observable, and/or partially-observable parameters.

Contributions

Herein, we consider the pure exploration problem for unstructured real-world environments. Our contribution is the use of the Poisson Exposure Process (Pep), which is extended from (Kim, Nefian, and Broxton 2010), to predict the availability of Kullback-Leibler (KL) divergence and to develop a bound for the sampling duration of uninformed exploration required before informed exploration can provide an exponential bound on the error between the actual and predicted KL divergence in unstructured real-world environments. Using the bound for the Pep, we develop an uninformed-to-informed exploration policy for unstructured real-world environments. Not only does the developed uninformed-to-informed exploration policy answer a relatively untreated question in the literature; i.e., at what point can an algorithm begin to exploit the available information, but the developed policy also demonstrates the human-like capability to capitalize on KL divergence, also called Bayesian surprise, as observed in (Itti and Baldi 2005).

Poisson Exposure Process Model

Similar to (Kim, Nefian, and Broxton 2010), we assume that our continuous observations (i.e., $v = D_{KL}(\hat{q}||\hat{p}) \sim \mathcal{V}$) are generated by an unknown monotonic transformation, $g(\cdot)$, of some draw from an unobserved discrete Poisson process.

$$\begin{aligned} y &\sim Pois(\lambda_y) \\ x &\sim \mathcal{X}|y \\ D_{KL}(\hat{q}||\hat{p}) &\sim Pep(\lambda|y), \end{aligned} \tag{1}$$

where λ_y is the arrival of unobserved informative events, $\lambda = g(y)$ and Pep is the Poisson exposure process. Hence, learning the Pep describing the available KL divergence allows us to learn about the underlying HMM, which we describe as a Poisson-arrival of entropy-injecting events.

Results

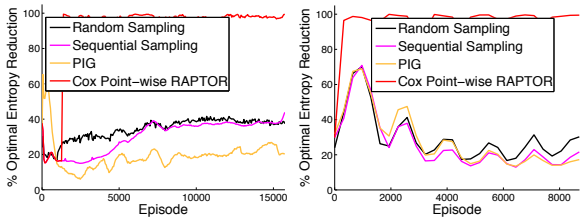
We test Algorithm 1 on the Intel-Berkeley temperature, European Research Area temperature, Washington rainfall, and Ireland wind-speed data sets in Figures 1a-1d. We leverage the Central Limit Theorem to treat the observations at each state as belonging to a Gaussian-distributed likelihood; i.e., $x_i \sim N(\mu, \sigma_L^2)$. For simplicity, each state is modeled

using the Gaussian distribution as the conjugate prior. The Bayesian update is

$$\mu_{\hat{q}} = \frac{\frac{\sigma^2_{L,i}}{M}\mu_{\hat{p}} + \sigma^2_{\hat{p},i}\bar{y}_i}{\frac{\sigma^2_{L,i}}{M} + \sigma^2_{\hat{p},i}} \text{ and } \sigma^2_{\hat{q}} = \left(\frac{\sigma^2_{L,i}}{M} + \sigma^2_{\hat{q}}\right)^{-1},$$

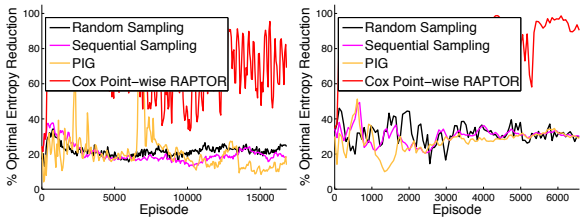
where $\mu_{\hat{p}}$ is the prior mean, $\mu_{\hat{q}}$ is the posterior mean, $\sigma^2_{\hat{p},i}$ is the prior variance, and $\sigma^2_{\hat{q},i}$ is the posterior variance. The KL divergence, D_{KL}, for scalar normal distributions (i.e., $d = 1$) is

$$D_{KL}(\hat{q}||\hat{p}) = \frac{log\left(\frac{\sigma^2_{\hat{p}}}{\sigma^2_{\hat{q}}}\right) + tr\left[\frac{\sigma^2{}_{\hat{q}}}{\sigma^2{}_{\hat{p}}}\right] - d + (\mu_{\hat{q}} - \mu_{\hat{p}})^T \sigma^{-2}_{\hat{p}} (\mu_{\hat{q}} - \mu_{\hat{p}})}{2}.$$



(a) Intel Berkeley Temperature Data Set Optimality (κ=6, K=52)

(b) European Temperature Data Set Optimality (κ=6, K=50)



(c) Washington Rainfall Data Set Optimality (κ=2, K=25)

(d) Ireland Wind-Speed Data Set Optimality (κ=2, K=12)

Figure 1: In the above subfigures, RAPTOR sequentially explores the data set until a domain exposure bound condition is satisfied. Once the bound is satisfied, RAPTOR has probabilistic guarantees on the prediction of Kullback-Leibler divergence, and the performance noticeably improves with respect to the baseline algorithms.

Conclusion

Herein, we present a bound for transitioning from an uninformed exploration policy to an informed exploration policy; this allows an agent, without preprocessing or a priori knowledge, to learn to exploit the task of exploration in partially-observable time-varying environments. The benefit of informed exploration policies which use information-theoretic quantities as the feedback signal is that exploration policy may minimize the amount of undiscovered information, given sufficient knowledge about the information dynamics, in an environment. However, informed exploration policies have traditionally either required preprocessing or a priori knowledge of the problem domain statistics, making them difficult to use in practice. Moreover, existing informed exploration policies such as IDS and PIG assume that the environment is stationary and observable, yet real-world environments are often time-varying and partially-observable. Unstructured real-world data experiments are used to validate the resultant uninformed-to-informed-exploration policy. The probability of best-action and the entropy reduction capability of the uninformed-to-informed-exploration policy exceeds that of baseline informed and uninformed exploration policies both in terms of the entropy reduction and the probability of selecting optimal actions.

Algorithm 1: RAPTOR with Pep-Cox Gaussian Process

Input: $t,\ \kappa,\ K,\ c,\ \delta,\ m,\ k$
Output: $V_i(\tau_i^{n_i}),\ \mathbb{E}(V_i(t)),\ \eta^t\ \forall\ i,\ t,\ n$
$(\alpha_i, \beta_i, n_i, \tau_i, \Delta t_i) = (0, 1, 0, 0, 0)\ \forall\ i$
For each: Epoch at time t
For each: $i \in S$
$\Delta t_i \leftarrow t - \tau_i^{n_i}\ \forall\ n_i \geq 1$
$\mathbb{E}[V_i|\Delta t_i] \leftarrow (1 - \sigma_i^2)(\mathbb{E}_{GP_i}(\Delta V_i|\Delta t_i) + V_i(\tau_i^{n_i})) + \sigma_i^2\ (\mathbb{E}_{Pep_i}(\Delta V_i|\Delta t_i) + V_i(\tau_i^{n_i}))$
End For
$b \leftarrow \underset{i}{\text{argmax}}\ \frac{1}{n_i \lambda_i (\tau_i^n - \tau_i^1)}$
If $t \geq \tau_b^n > \tau_b^1 + \frac{1}{nc\lambda}$ (Domain Exposure Bound)
$\eta^t \leftarrow \underset{\eta \subset \{1,\ldots,K\}}{\text{argmax}}\ \mathbb{E}\left[\sum_{i \in \eta} V_i\right]$
subject to $\text{Car}(\eta) \leq \kappa$
Else:
$\eta^t \leftarrow$Uninformed Policy (Sequential, Random, etc.)
End If
$n_i \leftarrow n_i + 1\ \forall\ i \in \eta_t$
Update: Pep-Cox Gaussian Process $\forall\ i \in \eta_t$
End For

Acknowledgements

This work is sponsored by the Department of Energy Award Number DE-FE0012173 and the Air Force Office of Scientific Research Award Number FA9550-14-1-0399.

References

Heyde, C., and Seneta, E. 1972. Studies in the history of probability and statistics. xxxi. the simple branching process, a turning point test and a fundamental inequality: A historical note on ij bienaymé. *Biometrika* 59(3):680–683.

Itti, L., and Baldi, P. F. 2005. Bayesian surprise attracts human attention. In *Advances in neural information processing systems*, 547–554.

Kim, T.; Nefian, A. V.; and Broxton, M. J. 2010. Photometric recovery of apollo metric imagery with lunar-lambertian reflectance. *Electronics letters* 46(9):631–633.

Little, D. Y., and Sommer, F. T. 2013. Learning and exploration in action-perception loops. *Frontiers in neural circuits* 7.

Russo, D., and Van Roy, B. 2014. Learning to optimize via information-directed sampling. In *Advances in Neural Information Processing Systems*, 1583–1591.