U.S. Department of Homeland Security
Centers of Excellence Summit 2018
University Research & Development to Protect the Homeland

Distributed Gaussian Process Regression for Efficient Robot Learning

John Martin, Doctoral Fellow
Mentor: Dr. Brendan Englot, Stevens Institute of Technology

Maritime Security Center — A Department of Homeland Security Center of Excellence
Stevens Institute of Technology — The Innovation University

Homeland Security Challenge

Context: Homeland security operations can sometimes be dangerous, replete with natural hazards of the environment, and sometimes necessitating extended services in locations where resources are sparse. The risk to humans, in many of these operations, can be reduced by automating tasks with robots.

Marine Robots: In this project, we focus on marine-based operations where robots can be employed. In the past, marine robots have been tasked to perform data gathering, inspect ship hulls, explore locations which may be too risky for humans to visit, serve as relays to increase network range, transport cargo, and create maps of uncharted territories.

Challenge: Robots require a lot of experience and computational resources to learn their tasks. Distributed algorithms can reduce the time it takes a robot to become proficient in a task. These algorithms have been developed previously for off-line applications, but currently no distributed methods exist for learning in a sequential setting.

Approach / Methodology

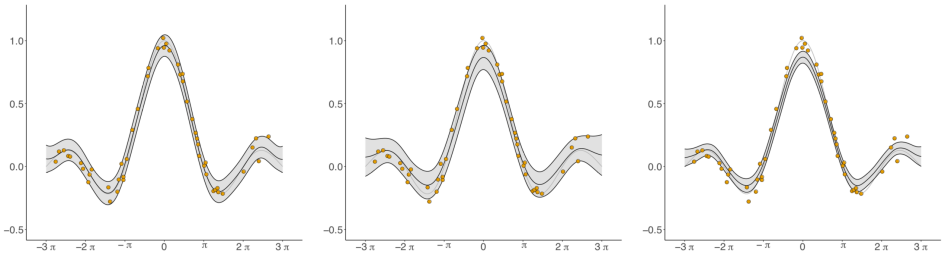
Methodology: We consider algorithms for Reinforcement Learning with Gaussian Process (GP) temporal differences. Our work studies the extent to which distributed computing can improve the amount of data GP-based value models can handle. By invoking episodic independence, we derive two different distributive models. One model represents the predictive value posterior as a sum of K experts, and the other, as a product. As such, predictions can be distributed to K independent processors. We propose actor-critic methods that exploit these models for efficient policy evaluation and action selection – balancing exploration and exploitation by maximizing the GP-UCB criterion. Our experiments compare the resulting methods to an actor-critic based on the standard GP Temporal Difference value model. We show our methods are able to process more data and, therefore, can solve complex problems which are too data-intensive for the standard model.

Reinforcement Learning: A learning paradigm in which robots select actions to maximize their value, $V(\mathbf{x}_n) = \mathbf{E}[D(\mathbf{x}_n)|R(\mathbf{x}_n), \mathbf{x}_n]$. Here, $D(\mathbf{x}) = V(\mathbf{x}) + \xi$, is the total of all randomized rewards, $R(\mathbf{x})$ the robot earns during operation. Rewards guide the robot to improve its actions, which is does by estimating $V(\mathbf{x})$through sequential observations.

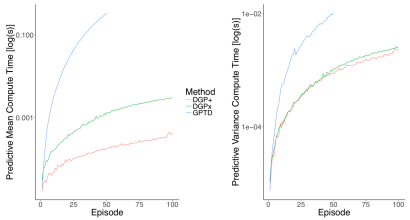
GP-Regression: As with simple regression and Kalman filtering, the goal is to estimate an unknown quantity. In our setting, we estimate the unknown value function $V(\mathbf{x})$. The statistical model is $R(\mathbf{x}) = V(\mathbf{x}_n) - \gamma V(\mathbf{x}_{n+1}) + \varepsilon_n$, where $\varepsilon_n = \xi_n - \gamma\xi_{n+1}$, $\xi \sim \mathcal{N}(0, \sigma^2)$.Stack values and rewards into vectors, $\mathbf{r}$, $\mathbf{v}$,assuming $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_{vv})$. The Gaussian Process assumes all variables are jointly Gaussian. To predict $V(\mathbf{x})$, we condition the joint model of V on $\mathbf{r}$, obtaining the posterior $\mathcal{N}(V(\mathbf{x})|v(\mathbf{x}), p(\mathbf{x}))$

$$v(\mathbf{x}) = \mathbf{k}_{r*}^\top(\mathbf{K}_{rr} + \mathbf{\Sigma})^{-1}\mathbf{r}, \qquad p(\mathbf{x}) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_{r*}^\top(\mathbf{K}_{rr} + \mathbf{\Sigma})^{-1}\mathbf{k}_{r*}.$$

Outcomes / Results



(a) GP (b) DGP+ (c) DGPx



(a) $v(\mathbf{x})$ Time (b) $p(\mathbf{x})$ Time

Two new methods: We approximate the standard model using a sum and product of distributions. We call these DGP+ and DGPx.

$$P(V|\mathbf{x}_*, \mathcal{D}) = \sum_{k=1}^{K} P_k(V|\mathbf{x}_*, \mathcal{D}_k).$$

$$P(V|\mathbf{x}_*, \mathcal{D}) = \prod_{k=1}^{K} P_k(V|\mathbf{x}_*, \mathcal{D}_k).$$

Take aways: Both algorithms approximate the standard method reasonably well. DGP+ is fastest, followed by DGPx, than the standard GP method.

Conclusion

To address a need for efficient robot learning methods, we developed two new algorithms for distributed reinforcement learning. In order to parallelize predictions, we approximated the standard model using independence assumptions. The distributed algorithms were shown to be reasonable approximations and to perform faster than the standard method. Future work aims to integrate these algorithms onto an underwater robot for physical learning trials.

Acknowledgements

"This material is based upon work supported by the U.S. Department of Homeland Security under Cooperative Agreement No. 2014-ST-061-ML0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security."