Learning Optimal Control Policies for an Autonomous Blimp with Non-stationary System Dynamics

Axel Rottmann Wolfram Burgard

***Abstract*— In this paper, we present an approach that applies the reinforcement learning principle based on Gaussian processes (GPs) to the problem of learning control policies for robotic blimps. In contrast to previous approaches, which often require specific and manually tuned a priori models, our method rather learns the policy online. Furthermore, we introduce efficient GP regression models which can deal with non-stationary underlying functions. This allows us to be highly adaptive to changing characteristics of the system during operation. In practical experiments carried out on a real robot we demonstrate that the system is able to learn a policy for height control online in less than one minute.**

I. Introduction

Compared to other flying vehicles, aerial blimps have the advantage that they operate at relatively low speed, that they do not need to move in order to keep their altitude, and additionally that they are not overly sensitive to control errors like, e.g., helicopters. In this paper, we investigate the problem of learning to control an autonomous blimp online without predefined physical models and under the assumption that the behavior of the system may change during runtime. Our approach is based on reinforcement learning and able to represent the system dynamics in a highly adaptive fashion to quickly react to situations when the behavior of the system changes. Furthermore, we continuously learn the value function according to this dynamics model to always obtain the optimal policy.

II. Online Learning

Reinforcement learning is based on the idea that an agent interacts with a potentially unknown environment and gets rewarded or penalized according to the actions it performs. Once we know the system dynamics of the blimp, the expected long term reward for each state and consequently the optimal policy can be learned using value evaluation. Therefore, our method learns both, the system dynamics and the value function.

A. System Dynamics

We learn the system dynamics iteratively by gathering transitions from state-action pairs while the blimp is in operation. This function is modeled with GPs, which are a powerful, non-parametric framework for regression [2]. Beside a predictive mean they also provide a predictive variance function. To be adaptive to a non-stationary behavior of the system, we present an extension to previous GP models that can appropriately handle outdated training samples by decreasing their influence on the predictive distribution [3]. The resulting model estimates for each sample of the training set an individual noise level and thereby detects outliers which will finally be removed.

Both authors are with the University of Freiburg, Department of Computer Science, D-79110 Freiburg, Germany

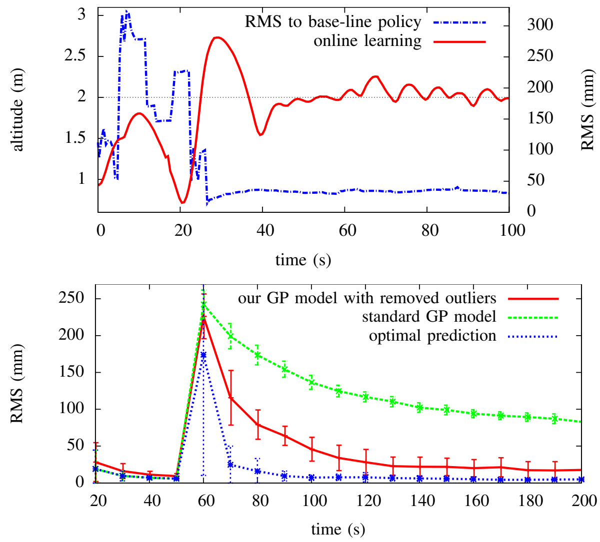

Fig. 1. Top: Altitude progress of the blimp while it learns to stabilize at 2 m and the corresponding error to a base-line policy. Bottom: Prediction accuracies of the system dynamics using different regression models when the behavior of the system has changed after 60 seconds.

B. Value Function

While learning the system dynamics, we constantly update the value function. As both functions are modeled with GPs, the uncertainty of the prediction of the behavior can be mapped efficiently onto the expected long term reward [1].

We performed several experiments on a real blimp to demonstrate the performance of our approach. A control policy to stabilize the altitude of the blimp was learned online in less than one minute (upper plot Fig. 1). Furthermore, as the system dynamics are modeled separately, our approach is able to quickly adapt the system dynamics and correspondingly the control policy to changed environmental conditions. In such situations, the presented GP extension outperforms previous regression models (lower plot Fig. 1).

References

[1] C. E. Rasmussen and M. Kuss, "Gaussian processes in reinforcement learning," in *Proc. of the Advances in Neural Information Proc. Systems 16 (NIPS)*. MIT Press, 2004.

[2] C. E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[3] A. Rottmann and W. Burgard, "Learning non-stationary system dynamics online using gaussian processes," in *DAGM-Symposium*, 2010, pp. 192–201.