Accepted Manuscript

Kernel dynamic policy programming: Applicable reinforcement learning to robot systems with high dimensional states

Yunduan Cui, Takamitsu Matsubara, Kenji Sugimoto

PII: S0893-6080(17)30143-0
DOI: http://dx.doi.org/10.1016/j.neunet.2017.06.007
Reference: NN 3772

To appear in: *Neural Networks*

Received date : 25 November 2016
Revised date : 8 May 2017
Accepted date : 19 June 2017

Kernel Dynamic Policy Programming: Applicable Reinforcement Learning to Robot Systems with High Dimensional States

Yunduan Cui*, Takamitsu Matsubara, Kenji Sugimoto

Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara, Japan

Abstract

We propose a new value function approach for model-free reinforcement learning in Markov decision processes involving high dimensional states that addresses the issues of brittleness and intractable computational complexity, therefore rendering the value function approach based reinforcement learning algorithms applicable to high dimensional systems. Our new algorithm, Kernel Dynamic Policy Programming (KDPP) smoothly updates the value function in accordance to the Kullback-Leibler divergence between current and updated policies. Stabilizing the learning in this manner enables the application of the kernel trick to value function approximation, which greatly reduces computational requirements for learning in high dimensional state spaces. The performance of KDPP against other kernel trick based value function approaches is first investigated in a simulated n DOF manipulator reaching task, where only KDPP efficiently learned a viable policy at $n = 40$. As an application to a real world high dimensional robot system, KDPP successfully learned the task of unscrewing a bottle cap via a Pneumatic Artificial Muscle (PAM) driven robotic hand with tactile sensors; a system with a state space of 32 dimensions, while given limited samples and with ordinary computing resources.

Keywords: Reinforcement learning, Kernel methods, Robot learning

1. Introduction

As one integral part of contemporary machine learning, reinforcement learning [1] enables agents to search for optimal policies by interacting with their environments without any prior knowledge and therefore becomes an approach that expresses a remarkably broad range of robot control problems in a natural manner [2]. Reinforcement learning is mainly divided into two groups:

*Corresponding author

Email addresses: cuiyunduan@gmail.com (Yunduan Cui), takam-m@is.naist.jp (Takamitsu Matsubara), kenji@is.naist.jp (Kenji Sugimoto)