REINFORCEMENT LEARNING USING GAUSSIAN PROCESSES FOR DISCRETELY CONTROLLED CONTINUOUS PROCESSES

M. DE PAULA and E.C. MARTÍNEZ

Instituto de desarrollo y diseño – INGAR (CONICET – UTN).
{marianodepaula, ecmarti}@santafe-conicet.gov.ar

Abstract— In many application domains such as autonomous avionics, power electronics and process systems engineering there exist discretely controlled continuous processes (DCCPs) which constitute a special subclass of hybrid dynamical systems. We introduce a novel simulation-based approach for DCCPs optimization under uncertainty using Reinforcement Learning with Gaussian Process models to learn the transitions dynamics descriptive of mode execution and an optimal switching policy for mode selection. Each mode implements a parameterized feedback control law until a stopping condition triggers. To deal with the size/dimension of the state space and a continuum of control mode parameters, Bayesian active learning is proposed using a utility function that trades off information content with policy improvement. Throughput maximization in a buffer tank subject to an uncertain schedule of several inflow discharges is used as case study addressing supply chain control in manufacturing systems.

Keywords— Hybrid Systems, stochastic systems, Optimization, Reinforcement Learning (RL), Gaussian Processes (GP).

I. INTRODUCTION

Modern automated systems are often constituted by interacting components of heterogeneous continuous/discrete nature. Dynamical systems having such a hybrid continuous/discrete nature are named hybrid systems (HS). We can find HS in electrical systems, chemical plants, biological systems, supply chains, unmanned vehicles, solar energy collectors, wind turbines and many others. A discretely controlled continuous process (DCCP) is a special type of hybrid systems where the discrete-event dynamics is the result of some event-based control strategy used to respond to external disturbances and endogenous inputs affecting the state evolution of the controlled system as whole (Simeonova, 2008; Goebel et al., 2009; Lunze and Lehmann, 2010).

A control strategy is implemented by a switching policy which timely stops an operating mode due to goal achievement, a state constraint or an external event (Mehta and Egerstedt, 2006). Each control mode, or simply “mode,” implements a parameterized feedback control law until a terminating condition is activated. Then, each mode differs from other by its parameterization which varies in a continuum. Duration time for each mode execution depends on the type of behavior or specific goal which is being pursued and occurrence of disturbances and events affecting the system dynamics. For optimal control, the switching policy must generate a sequence of control modes to complete a goal-directed control task from different initial states while minimizing some performance criterion (Görges et al., 2011). Thus, optimal operation of a DCCP gives room for resorting to a Lebesgue sampling strategy of states to advantage (Xu and Cao, 2011). This paper deals with the problem of finding a switching policy for optimal operation of a DCCP under uncertainty so as to implement a goal-directed control strategy in real-time.

For a finite number of modes, a novel modeling paradigm known as integral continuous-time hybrid automata (icHA) has been recently proposed for event-driven optimization-based control for which no a priori information about the timing and order of the events is assumed (Di Cairano et al., 2009). The solution of dynamic optimization problems with continuous time hybrid automata embedded has been thoroughly reviewed by Barton et al. (2006). In a more recent work, approximate dynamic programming has been successfully applied to the discrete-time switched LQR control problem (Zhang et al., 2009). The important issue of optimality in multi-modal optimal control has also been addressed by Mehta and Egerstedt using reinforcement learning (RL) techniques regarding a finite set of control modes (Mehta and Egerstedt, 2006).

Uncertainty in the initial states is a major obstacle for multi-modal control since fixed control programs are derived from assumed initial conditions. Reinforcement Learning (RL) (Sutton and Barto, 1998) is a simulation-based approach to solve optimal control problems under uncertainty. For optimal operation of DCCPs under uncertainty a multi-modal control program should be able to adapt on-line in order to handle disturbances or events that may severely affect a control program performance or renders it even unfeasible. In this work, the main argument is that for optimal operation of a DCCP under uncertainty a switching policy is required to implement goal-directed control in real-time. A novel simulation-based algorithm which combines dynamic programming with Lebesgue sampling and Gaussian process (Rasmussen and Williams, 2006) approximation is proposed to learn a switching policy for mode selection. To deal with the size and dimension of the state space and a continuum of feedback law parameters, Bayesian active learning (Deisenroth et al., 2009) is proposed using a utility function that trades off information content with switching policy improvement. Probabilistic models of the state transition dynamics following each mode execution are learned upon data obtained by increasing biasing operating conditions.