Neurocomputing 156 (2015) 157–165

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Optimal control of nonlinear discrete time-varying systems using a new neural network approximation structure

Bahare Kiumarsi [a,*], Frank L. Lewis [b,c], Daniel S. Levine [d]

[a] UTA Research Institute, University of Texas at Arlington, Fort Worth, TX 76118, USA
[b] UTA Research Institute, University of Texas at Arlington, Fort Worth, TX, USA
[c] State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang, China
[d] University of Texas at Arlington, Arlington, TX, USA

ARTICLE INFO

Article history:
Received 4 September 2014
Received in revised form
4 December 2014
Accepted 24 December 2014
Communicated by: D. Liu
Available online 6 January 2015

Keywords:
Multiple-model
Optimal control
Adaptive self- organizing map
Reinforcement learning
Value function approximation

ABSTRACT

In this paper motivated by recently discovered neurocognitive models of mechanisms in the brain, a new reinforcement learning (RL) method is presented based on a novel critic neural network (NN) structure to solve the optimal tracking problem of a nonlinear discrete time-varying system in an online manner. A multiple-model approach combined with an adaptive self-organizing map (ASOM) neural network is used to detect changes in the dynamics of the system. The number of sub-models is determined adaptively and grows once a mismatch between the stored sub-models and the new data is detected. By using the ASOM neural network, a novel value function approximation (VFA) scheme is presented. Each sub-model contributes into the value function based on a responsibility signal obtained by the ASOM. The responsibility signal determines how much each sub-model contributes to the general value function. Novel policy iteration and the value iteration algorithms are presented to find the optimal control for the partially-unknown nonlinear discrete time-varying systems in an online manner. Simulation results demonstrate the effectiveness of the proposed control scheme.

1. Introduction

Design of feedback controllers for unknown and uncertain systems has been widely considered in the control system community due to its potential applications in industry. Accordingly, numerous adaptive and robust control methods have been proposed to deal with system uncertainties. Neural networks (NNs) have been extensively used to develop stable and robust adaptive controllers to compensate for uncertainties and disturbances. However, most existing work in the literature concerns only the stability of adaptive NN controllers and not other performance criteria. Recently, a great deal of interest in developing adaptive optimal controller using NNs has been witnessed [1–6]. These methods learn the optimal solution successfully for both known and unknown, but time-invariant systems. However, in many real world situations the controlled system dynamics are subjected to abrupt changes caused by component failures or repairs, changing sub-model interconnections, or sudden changes in environmental factors, etc.

In past decades, multiple-model methods have been presented for control of complex and/or time-varying systems. One approach of the multiple-model methods assumes that a nonlinear process with a high degree of uncertainty is represented by a collection of linear models with a low degree of uncertainty. Another approach to the multiple-model is that, for a system under abrupt changes in dynamics, several models are considered and each model belongs to a distinct subset of the system behavioral space. In multiple-model approaches, a controller is assigned corresponding to each sub-model, then at each time, an active controller is selected using a supervisory unit (which is usually an estimator-based supervisor) using logical decision rules. The supervisory either switches between the different sub-model controllers or combine them together based on certain criteria. Both fixed and adaptive models have been used for control and identification of time-varying dynamical systems [7–13]. Existing multiple-model approaches can guarantee stability of the system dynamics but they are generally far from optimal.

Reinforcement learning (RL) refers to a class of methods that enable the design of adaptive optimal controllers for uncertain dynamical systems. RL methods learn online, in real time, the solutions to user-prescribed optimal control problems for uncertain systems using only measured data from the controlled system [14–20]. RL mechanisms operate in the human brain, where the dopamine neurotransmitter in the basal ganglia acts as a

* Corresponding author at: 7300 Jack Newell Blvd S, Fort Worth, TX, USA, 76118. Tel.: +6822410845.
E-mail address: kiumarsi@uta.edu (B. Kiumarsi).

http://dx.doi.org/10.1016/j.neucom.2014.12.067
0925-2312/

reinforcement informational signal that provides learning at the level of the neuron [21]. One type of reinforcement learning algorithm employs the actor-critic structure [22]. This structure produces forward-in-time algorithms that are implemented in real time where in an actor component applies an action, or control policy, to the environment, and a critic component assesses the value of that action. The learning mechanism supported by the actor-critic structure has two steps, namely, policy evaluation, executed by the critic, followed by policy improvement, performed by the actor.

The standard RL algorithms cannot be used for time–varying systems with abrupt changes in dynamics. To overcome this issue, in [23], an RL architecture based on multiple-model approach is presented. In this multiple-model approach, each sub-model consists of a state prediction model and an actor-critic controller. However, in the method of [23], it is implicitly assumed that the number of sub-models is known a priori, which may not be realistic. Moreover, state derivatives prediction errors are used to classify sub-models. This requires identifying of the system dynamics in each sub-model.

Recent research in neurocognitive psychology shows that there are several actor-critic structures in the brain [23–25]. The actor-critic structure of basal ganglia, cerebellum, and muscle motor control is well known, with dopamine neurons in the basal ganglia providing the RL mechanism. Humans naturally trade off open-loop control based on stored memories and closed-loop control based on real-time learning for different tasks and spatial contexts. Emotion and past stored experiences can significantly affect decision-making in humans. The work in [26–28] details how stored behavior patterns can be used to enact fast decisions by drawing on previous experiences when there is a match between observed attributes and stored patterns. In the event of risk, mismatch, or anxiety, higher-level control mechanisms in the brain are recruited that involve more deliberative actions based on refinements in current real-time observations.

There are several interactions between relevant brain regions. First, in the amygdala and orbital frontal cortex (OFC) interaction, the amygdala seems to play a role in determining which aspects of a problem are relevant for decisions. If the match of a problem with previously seen situations is close, decision activity occurs primarily in the amygdala-OFC system. Second, in the interaction of the anterior cingulate cortex (ACC) and dorsolateral prefrontal cortex (DLPFC), if mismatch occurs, decision involving higher-order cognition is required. This involves recruiting higher-level brain structures including the ACC and DLPFC. In these activities, real-time information from the environment is more closely examined to construct closed-loop feedback controls that perform better with guaranteed results. ACC and DLPFC seem to be recruited in emotionally undesirable decisions that do not match previously stored successful situations or have increased complexity. These regions play a role in selective attention, dynamic focusing of awareness onto the details in a problem, and in deliberative cognition [29–31].

Self-organizing map (SOM) is a very well established method in the field of neural networks [32,33]. The SOM has been used in many industrial applications. An important application of SOM is clustering, which attempt to organize unlabeled input vectors into clusters or "natural groups" such that points within a cluster are more similar to each other than vectors belonging to different clusters. The main shortcoming of the SOM is that it requires either specifying the number of clusters in advance. To overcome this restriction, adaptive SOM algorithm has been proposed.

Motivated by recent neurocognitive models of mechanisms in the brain, a new interacting structure of RL approach is presented in this paper to find the optimal policy for a time-varying system without requiring to have or to identify the knowledge of the drift system dynamics. The number of the sub-models is not assumed to be fixed, and is growing adaptively based on some criteria. A multiple model approach combined with adaptive self-organizing map (ASOM) network is used to detect the change on the dynamics of the system. The ASOM operates similar to the amygdala-OFC and the ACC-DLPFC in the brain. The number of sub-models is determined adaptively and grows once a mismatch between the stored sub-models and the new data is detected. That is, once the performance of the existing sub-models is not satisfactory, a new sub-model is needed. By using the ASOM neural network, a new structure for approximating the optimal value function of the discrete time-varying systems is presented. Each sub-model contributes into the value function based on a responsibility signal obtained by the ASOM. The responsibility signal of each sub-model at each time is an indication of how likely the current input is to be generated by the sub-models at that time. Based on the presented value function, policy iteration and value iteration algorithms are presented to find the optimal control for the discrete time-varying systems. The knowledge of the drift system dynamics of sub-models is not required to be known or to be identified and the optimal control is obtained using measured data in real time.

The contributions of the paper are as follows.

1. Motivated by neurocognitive models of amygdala-OFC mechanisms in the brain, an adaptive self-organizing map (ASOM) neural network is used to cluster input-output data into different regions, each belonging to different sub-models. The number of sub-models is assumed unknown. New sub-models are built incrementally, in real time.
2. A new structure for approximation and updating of the value function is presented for discrete time-varying systems that is based on different sub-models to allow the use of previous experience in fast response to environmental cues. In this structure, the value function approximation (VFA) can be realized through cooperative learning in the sense that each sub-model learns not only from data samples from its corresponding region, but also from samples from its neighboring regions.
3. We present a partially model free method for discrete time-varying systems based on reinforcement learning. The proposed method does not require identifying or complete dynamics of the sub-models. It finds the optimal solution using only measured data.

This paper is organized as follows. Optimal tracking control for nonlinear time-varying systems is presented in Section 2. New formulation for the nonlinear tracking problem is presented in Section 3. New value function structure and adaptive self-organizing map neural network for the nonlinear discrete time-varying systems are proposed in Section 4. In Section 5, online policy iteration and value iteration algorithms are developed to solve the optimal control problem for partially unknown nonlinear discrete time-varying systems. Simulation results of the mentioned algorithms are presented in Section 6.

2. Optimal tracking control for nonlinear time-varying systems

In this section the optimal tracking problem for nonlinear discrete time-varying systems is defined. Consider the nonlinear affine discrete time-varying system given by

$$x(k+1)=f(x(k),k)+g(x(k))\,u(k) \tag{1}$$

where $x(k)\in R^n$ represents the state vector of the system, $u(k)\in R^m$ represents the control vector, $f(x(k),k)\in R^n$ is the drift dynamics of the system, and $g(x(k))\in R^{n\times m}$ is the input dynamics of the system. Assume that $f(0)=0$ and $f(x)+g(x)u$ is Lipschitz continuous on a compact set Ω which contains the origin, and the system (1) is controllable in the sense that there exists a continuous control on Ω which stabilizes the system. The drift dynamics is assumed to be under