Neurocomputing 149 (2015) 116–123

Contents lists available at ScienceDirect

Neurocomputing

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

A neural network based online learning and control approach for Markov jump systems

Xiangnan Zhong [a], Haibo He [a,*], Huaguang Zhang [b], Zhanshan Wang [b]

[a] Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, Kingston, RI 02881, USA
[b] School of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004, China

ARTICLE INFO

Article history:
Received 11 June 2013
Received in revised form
7 October 2013
Accepted 14 January 2014
Available online 12 August 2014

Keywords:
Adaptive dynamic programming
Neural network
Markov jump systems
Optimal control

ABSTRACT

In this paper, we propose an optimal online control method for discrete-time nonlinear Markov jump systems (MJSs). The Markov chain and the weighted sum technique are introduced to convert the Markov jumping problem into an optimal control problem. We then use adaptive dynamic programming (ADP) to accomplish online learning and control with specific learning algorithm and detailed stability analysis, including the convergence of the performance index function sequence and the existence of the corresponding admissible control input. Neural networks are applied to implement this ADP approach and online learning method is used to tune the weights of the critic and the action networks. Two different numerical examples are given to demonstrate the effectiveness of the proposed method.

1. Introduction

Adaptive dynamic programming (ADP) has been widely recognized as one of the "core methodologies" to achieve optimal control in stochastic process in a general case to achieve intelligent control [1,2]. Taking the advantage of approximating the solutions of optimal control problems in equivalent to solve the Hamilton–Jacobi–Bellman (HJB) equation, this method has attracted significantly increasing attention in recent years. Extensive efforts and promising results in both theoretical research and engineering applications have been achieved over the past decades. Among these achievements, we highlight Al-Tamimi et al. [3], Abu-Khalaf et al. [4], Wei and Liu [5–7], Lewis and Vamvoudakis [8,9], He et al. [10–12], Zhang et al. [13–15], Si et al. [16–18], He and Jagannathan [19–21], Seiffertt et al. [22], Zhong et al. [23,24] and Lin et al. [25,26] from the theoretical perspective that are closely related to the research presented in this paper. These achievements cover a large variety of problems, including system stability, convergence proof, optimal control, and state prediction. Interested readers can refer to the two important handbooks [27,28] on ADP for many other successful architectures, algorithms and challenging engineering applications.

On the other hand, there has been extensive interest in the stability analysis and controller design of the Markov jump systems (MJSs) over the past decades due to its powerful modeling capability for power systems [29,30], aerospace systems [31], and manufacturing systems [32–34]. In practice, random parameters change may exist in these systems resulting from sudden environmental disturbances, abrupt changes of the operating point, or component failure or repairs. These make the systems that cannot be easily modeled. The studies of MJSs build a bridge between these architecture systems and the theoretical analysis. However, many of the research in this field completely depend on the accuracy system functions [35–39], which narrows the range of application of this powerful modeling method. How to solve the problems of MJSs without the knowledge of system functions is a challenging topic.

This paper proposes an optimal control method for a class of discrete-time nonlinear Markov jump systems without the requirement of system functions by using ADP technique. Note that, generally, ADP can be categorized into three typical structures which are heuristic dynamic programming (HDP), dual heuristic dynamic programming (DHP), and globalized dual heuristic dynamic programming (GDHP). This paper is focused on the HDP technique for stabilizing the MJSs. The main contribution of this work is to introduce the ADP method into the field of MJSs by transforming the MJSs control problem with multi-subsystem into a single-objective optimal control problem. Moreover, unlike the traditional method to solve the MJSs problem, such as linear matrix inequality, our approach based on ADP technique is an adaptive and learning process, which means when the parameters

* Corresponding author.
E-mail addresses: xzhong@ele.uri.edu (X. Zhong), he@ele.uri.edu (H. He), hgzhang@ieee.org (H. Zhang), zhanshan_wang@163.com (Z. Wang).

http://dx.doi.org/10.1016/j.neucom.2014.01.060
0925-2312/

of the systems are changed, our approach can still find the optimal controller adaptively. Besides, the convergence of the proposed ADP approach is provided in detail. Two numerical examples are presented to verify the validity of the proposed method.

The rest of this paper is organized as follows. In Section 2, we formulate the MJSs control problem analyzed in this paper. The performance index function for the whole MJSs is obtained by combining the performance index functions for the subsystems using Markov chain and weighted sum technique. The ADP method for discrete-time nonlinear MJSs is established in Section 3. Section 4 presents the detailed convergence analysis of the performance index function and the existence of the admissible control for the proposed optimal control scheme. Neural networks are used in Section 5 to implement this ADP scheme. The online learning method is also applied in this section to tune the weights of the critic and the action networks. In Section 6, two numerical examples, including one with two jumping modes and one with four jumping modes, are presented to demonstrate the effectiveness of the proposed approach. Finally, Section 7 concludes the paper.

2. Problem statement

The discrete-time nonlinear MJSs can be described by the following equation:

$$x(k+1)=F(x(k),u(k),\theta(k)), \quad k\geq 0 \tag{1}$$

where $x(k)\in R^n$ is the system state with the initial state $x(0)$ and $u(k)\in R^m$ is the control vector. $F(x(k),u(k),\theta(k))$ denotes the unknown system function and $F(0,0,\theta(k))=0$. We assume that $F(x(k),u(k),\theta(k))$ is Lipschitz continuous. $\{\theta(k),k\geq 0\}$ is the discrete-time Markov chain, which refers to the active mode of the whole system in each step and takes values in a finite set $S=\{1,2,\ldots,l\}$, where l is the number of the subsystems. The elements in the Markov chain are given by

$$p_{ij}=\text{Prob}(\theta(k+1)=j|\theta(k)=i) \tag{2}$$

which denotes the transition probability that the next active subsystem is the jth one given that the current active subsystem is the ith one. Hence, we know $p_{ij}\geq 0,\ \forall i,j\in S$ and $\sum_{j=1}^{l}p_{ij}=1$.

Define the performance index function for each subsystem as follows:

$$J(x(k),\theta(k))=\sum_{t=k}^{\infty}\alpha^{t-k}U(x(t),u(t),\theta(t)) \tag{3}$$

where the utility function $U(x(t),u(t),\theta(t))=Q(x(t),\theta(t))+u(t)^T R(\theta(t))u(t)$ is positive definite, i.e., $U(x(t),u(t),\theta(t))=0$, if and only if $x(t)=0$ and $u(t)=0$; otherwise $U(x(t),u(t),\theta(t))>0$. And, $0<\alpha\leq 1$ is the discount factor.

In the following part, we use $F_i(x(k),u(k))$, $J_i(x(k))$, $U_i(x(k),u(k))$, $Q_i(x(k))$, R_i to represent the notation $F(x(k),u(k),\theta(k))$, $J(x(k),\theta(k))$, $U(x(k),u(k),\theta(k))$, $Q(x(k),\theta(k))$, $R(\theta(k))$ to simplify the presentation.

For optimal control problem, it is desired to find an optimal control $u^*(k)$ to minimize the performance index function for system (1). However, due to the existence of the transition probabilities (2), we cannot just add all the performance index functions of the subsystems to act as that of the whole MJSs. Here, we use

$$\begin{cases} J_I(x(k))=p_{11}J_1(x(k))+p_{12}J_2(x(k))+\cdots+p_{1l}J_l(x(k)) \\ J_{II}(x(k))=p_{21}J_1(x(k))+p_{22}J_2(x(k))+\cdots+p_{2l}J_l(x(k)) \\ \vdots \\ J_L(x(k))=p_{l1}J_1(x(k))+p_{l2}J_2(x(k))+\cdots+p_{ll}J_l(x(k)) \end{cases} \tag{4}$$

to reconstruct the performance index function according to the Markov chain (2). Then, by using the weighted sum technique, the final performance index function for MJSs is obtained as

$$J(x(k))=\omega_1 J_I(x(k))+\omega_2 J_{II}(x(k))+\cdots+\omega_l J_L(x(k)) \tag{5}$$

where $\omega_i>0$ is the weight vector and $\sum_{i=1}^{l}\omega_i=1$.

Hence, the control vector $u(k)$ needs to be found to minimize the performance index function (5) and make the MJSs achieve stability. Note that this control law must not only stabilize the system on the compact set $\Omega\in R^n$, but also guarantee that (5) is finite, which is called admissible control.

Definition 1. A control law is said to be an admissible control with respect to (5) on Ω, if $u(k)$ is continuous on Ω and can stabilize system (1) for all $x(0)\in\Omega$, $u(k)=0$ as $x(k)=0$, and for $\forall x(k)$, $J(x(k))$ is finite.

3. ADP approach for optimal control problem of nonlinear MJSs

In this section, the ADP approach for discrete-time nonlinear MJSs is presented based on the performance index function (5).

Eq. (5) can be expanded as

$$\begin{aligned} J(x(k)) &= \omega_1 J_I(x(k))+\omega_2 J_{II}(x(k))+\cdots+\omega_l J_L(x(k)) \\ &= (\omega_1 p_{11}+\omega_2 p_{21}+\cdots+\omega_l p_{l1})J_1(x(k)) \\ &\quad +(\omega_1 p_{12}+\omega_2 p_{22}+\cdots+\omega_l p_{l2})J_2(x(k)) \\ &\quad +\cdots+(\omega_1 p_{1l}+\omega_2 p_{2l}+\cdots+\omega_l \pi_{ll})J_l(x(k)) \\ &= D_1 J_1(x(k))+D_2 J_2(x(k))+\cdots+D_l J_l(x(k)) \\ &= \sum_{i=1}^{l}\sum_{t=k}^{\infty}(\alpha^{t-k}D_i U_i(x(t),u(t))) \end{aligned} \tag{6}$$

where $D_i=\sum_{j=1}^{l}\omega_j p_{ji}>0$. Hence, Eq. (6) is positive definite, i.e., the above performance index function serves as a Lyapunov function.

The equivalent equation of (6) is given by the Bellman equation:

$$\begin{aligned} J(x(k)) &= \sum_{i=1}^{l}(D_i U_i(x(k),u(k)))+\sum_{i=1}^{l}\sum_{t=k+1}^{\infty}\alpha^{t-k}D_i U_i(x(t),u(t)) \\ &= \sum_{i=1}^{l}(D_i U_i(x(k),u(k)))+\alpha\sum_{i=1}^{l}\sum_{t=k+1}^{\infty}\alpha^{t-(k+1)}D_i U_i(x(t),u(t)) \\ &= D^T U(x(k),u(k))+\alpha J(x(k+1)) \end{aligned} \tag{7}$$

where
$D=(D_1,D_2,\ldots,D_l)^T$,
$U(x(k),u(k))=(U_1(x(k),u(k)),U_2(x(k),u(k)),\ldots,U_l(x(k),u(k)))^T$.

Depending on Bellman's optimality principle, the optimal performance index function $J^*(x(k))$ is time invariant and satisfies the discrete-time HJB equation:

$$J^*(x(k))=\min_{u(k)}\{D^T U(x(k),u(k))+\alpha J^*(x(k+1))\} \tag{8}$$

Besides, the optimal control $u^*(k)$ satisfies the first-order necessary condition, which is obtained by gradient of the right-hand side of (8) with respect to $u(k)$ as

$$\frac{\partial(D^T U(x(k),u(k)))}{\partial u(k)}+\alpha\left(\frac{\partial x(k+1)}{\partial u(k)}\right)^T\frac{\partial J^*(x(k+1))}{\partial x(k+1)}=0 \tag{9}$$

Therefore, the optimal control policy can be expressed as

$$u^*(k)=-\frac{\alpha}{2}\left(\sum_{i=1}^{m}D_i R_i\right)^{-1}\left(\frac{\partial F_i(x(k),u(k))}{\partial u(k)}\right)^T\frac{\partial J^*(x(k+1))}{\partial x(k+1)} \tag{10}$$

where $J^*(x(k))$ is solved in the following HJB equation:

$$J^*(x(k))=\sum_{i=1}^{m}D_i Q_i(x(k))+\frac{\alpha^2}{4}\left(\left(\frac{\partial F_i(x(k),u(k))}{\partial u(k)}\right)^T\frac{\partial J^*(x(k+1))}{\partial x(k+1)}\right)^T$$