Neurocomputing 239 (2017) 1–8

Contents lists available at ScienceDirect

Neurocomputing

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

Neural-network-based control scheme for a class of nonlinear systems with actuator faults via data-driven reinforcement learning method

He Jiang, Huaguang Zhang*, Yang Liu, Ji Han

College of Information Science and Engineering, Northeastern University, P.O. Box 134, 110819 Shenyang, PR China

A R T I C L E I N F O

Article history:
Received 16 May 2016
Revised 15 January 2017
Accepted 18 January 2017
Available online 27 January 2017

Communicated by Ding Wang

Keywords:
Reinforcement learning
Adaptive dynamic programming
Data-driven
Model-free
Neural networks

A B S T R A C T

This paper investigates the fault tolerant control problem for a class of continuous-time nonlinear systems with completely unknown dynamics via the data-based adaptive dynamic programming method. The proposed controller can be divided into two parts: (1) optimal control policy of the fault-free systems and (2) fault compensator. Firstly, a model-based policy iteration algorithm is introduced to obtain the optimal control law. Subsequently, a fault compensator is derived to get rid of the impact of the actuator fault. The stability analysis of the model-based control scheme is presented by using Lyapunov theory. However, for the complex practical systems, system models are generally unavailable, and thus the model-based approaches may be invalid. To overcome this difficulty, we provide a data-driven reinforcement learning method and an identification approach to design the two parts of the proposed controller, respectively, without any knowledge of the system models. Neural networks are employed to implement these two data-based methods. Finally, two simulation examples are shown in details to demonstrate the effectiveness of our proposed scheme.

1. Introduction

In the past few years, industrial systems have become increasingly expensive and complex, such as power systems, manufacturing systems, communication systems and aero systems, where a higher demand of system safety and reliability is necessary. However, as the complexity of system structure and the amount of system components increase, diversities of unexpected faults are inevitable. Faults may cause the deviation of system parameters or characteristic properties from the standard level, or even break down the whole system. Among all types of possible faults, the actuator ones are regarded as one of the most interesting and challenging research topics, for the control performance degradation and serious damages may be directly induced by the wrong actuator actions.

In order to obtain higher system reliability and safety, and better control performance, many significant works concerning the fault tolerant control (FTC) have been proposed to handle the actuator faults. In the work [1], a novel FTC scheme was developed for a class of nonlinear systems with the unknown time-varying fault and the dead-zone via a nonlinear fault estimator, and then this method was further extended to solve the fault tolerant tracking control problem through the backstepping technique. Authors of [2] applied an uncertainty decomposition based adaptive FTC method to deal with the attitude tracking problem of flexible spacecraft with unknown actuator faults and inertia parameters. By using a new local fault identification algorithm and a global adaptive sliding-mode observer, a decentralized fault tolerant controller was proposed for near-space vehicle attitude dynamics with different types of actuator faults in [3]. Authors of [4] utilized the implicit function theorem, and designed a novel neural network (NN) based active FTC approach with the fault alarm. In [5], a fuzzy adaptive control scheme was presented to solve the consensus problem of a class of networked unknown nonlinear multi-agent systems with time-varying actuator faults. Unfortunately, for the most real-world large-scale control systems, the system structures are generally so complex that the accurate system mathematical models are unavailable to support the control decision design.

Reinforcement learning (RL) [6–11], which is inspired by the idea of machine learning and computational intelligence, has developed many useful ways to let an agent find out the optimal action through the responses from its environment, such as adaptive/approximate dynamic programming (ADP) [12,13]. ADP methods can be generally classified into five main branches, which include heuristic dynamic programming (HDP), dual HDP (DHP), globalized DHP (GDHP), action-dependent HDP (ADHDP),

* Corresponding author.
E-mail addresses: jianghescholar@163.com (H. Jiang), hgzhang@ieee.org, zhg516516@gmail.com (H. Zhang), yangliu9611@163.com (Y. Liu), hanji0912@163.com (J. Han).

http://dx.doi.org/10.1016/j.neucom.2017.01.047
0925-2312/

and action-dependent DHP (ADDHP). There are also a variety of iterative algorithms within the scope of ADP, which can be divided into two mainstream iterative ones, namely, policy iteration (PI) [14–16] and value iteration (VI) [17,18]. Based on the theoretical structure constructed by these classical works [14–18], various optimal control problems have been addressed, such as robust optimal control [19,20], optimal tracking control [21–25], optimal control with time-delay [26–28] and constrained control input [29–33], optimal control for zero-sum [34–39] and non-zero-sum games [40–43], and optimal control applied on multi-agent systems [44–48] and unknown systems [49–52].

Recently, researchers have started to apply RL and ADP methods to address the FTC issues. In [53], authors studied the FTC problem by using globalized dual heuristic programming through an online learning supervisor, and then, this approach was developed and improved in [54]. A neural-network-based adaptive actuator fault compensation control scheme was proposed for a class of uncertain discrete-time systems with triangular forms in [55]. Authors of [56] utilized a RL method to solve the FTC problem of a class of multiple-input-multiple-output nonlinear systems. However, among most relevant research results, the knowledge of accurate system models is always required. To the best of our knowledge, there are still few works concerning the model-free FTC issues by using RL approaches, which motivates our research work of this paper.

In this paper, a novel neural-network-based control scheme is designed for a class of nonlinear systems with completely unknown dynamics and actuator faults via data-driven ADP methods. Firstly, the problem statement is given in Section 2. Secondly, a fault compensator based controller is derived via PI algorithm, and the stability analysis is also provided through Lyapunov theory in Section 3. Thirdly, a data-driven RL method and an identification scheme, implemented by NNs, are employed to obtain the two parts, which constitute the proposed controller, respectively, without any knowledge of system models in Section 4. Section 5 provides two simulation examples to demonstrate the effectiveness of our proposed scheme. Finally, a brief conclusion is given in Section 6.

2. Problem statement

Consider a class of nonlinear systems with actuator faults as follows:

$$\dot{x}(t) = f(x(t)) + g(x(t))(u(t) - \Gamma) \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state; $u(t) \in \mathbb{R}^m$ is the control input; $\Gamma \in \mathbb{R}^m$ denotes the unknown constant actuator fault, which implies $\dot{\Gamma} = 0$; $f(x(t)) \in \mathbb{R}^n$ and $g(x(t)) \in \mathbb{R}^{n\times m}$ represent the system matrices, which are both considered to be unknown in this paper.

Assumption 1 [16]**.** $f(\cdot)$ is locally Lipchitz and differentiable in its arguments with $f(0) = 0$, i.e., there exists a positive constant b_f such that $\|f(x)\| < b_f\|x\|$. $g(\cdot)$ is bounded by a constant b_g, i.e., $\|g(x)\| < b_g$.

For the system (1) with the fault-free condition, i.e., $\Gamma = 0$, we can define the performance index function as:

$$V(x(t)) = \int_t^\infty (x^T(\tau)Qx(\tau) + u^T(\tau)Ru(\tau))d\tau \tag{2}$$

where $Q \in \mathbb{R}^{n\times n} > 0$ and $R \in \mathbb{R}^{m\times m} > 0$ are both positive definite symmetric matrices.

Definition 1 [30]**.** (Admissible Control) A state feedback control policy $u(x)$ for the system (1) with the condition $\Gamma = 0$ is said to be admissible with respect to $V(x)$ on a set $\Omega \subseteq \mathbb{R}^n$, if $u(x) \in \Psi(\Omega)$ not only stabilizes system (1) but also guarantees $V(x)$ finite.

According to Leibniz's rule, differentiating $V(x)$ along the system (1) with $\Gamma = 0$ yields

$$0 = \nabla V^T(x)(f(x) + g(x)u) + x^TQx + u^TRu \tag{3}$$

where $\nabla V(x) \triangleq \partial V(x)/\partial x$.

Based on (3), the Hamiltonian function can be defined as

$$H(x, u, \nabla V(x)) \triangleq \nabla V^T(x)(f(x) + g(x)u) + x^TQx + u^TRu. \tag{4}$$

The optimal performance index function $V^*(x)$ can be expressed as

$$V^*(x(t)) = \min_{u\in\Psi(\Omega)} \int_t^\infty (x^T(\tau)Qx(\tau) + u^T(\tau)Ru(\tau))d\tau \tag{5}$$

which also satisfies the Hamilton–Jacobi–Bellman (HJB) equation as below:

$$0 = \min_{u\in\Psi(\Omega)} H(x, u, \nabla V^*(x)). \tag{6}$$

If the minimum on the right hand side of (6) exists and is unique, then the optimal control policy $u^*(x)$ can be given by

$$u^*(x) = -\frac{1}{2}R^{-1}g^T(x)\nabla V^*(x). \tag{7}$$

Substituting (7) into (6), the HJB equation is rewritten as

$$\nabla V^{*T}(x)f(x) + x^TQx - \frac{1}{4}\nabla V^{*T}(x)g(x)R^{-1}g^T(x)\nabla V^*(x) = 0. \tag{8}$$

Remark 1. For the linear systems, the HJB equation can be reduced to the well-known Riccati equation, which can be solved directly. Nevertheless, for the nonlinear systems, the HJB equation is a nonlinear partial differential one, which is difficult or impossible to obtain the solution analytically. To overcome this difficulty, the PI algorithm will be introduced in the following section.

3. Model-based FTC scheme via PI algorithm

In this section, we firstly present the PI algorithm to achieve the optimal control policy when there is no additive actuator fault, and then, a fault compensator is added to accomplish the design of the controller.

3.1. Model-based PI algorithm

The PI algorithm [12,14,57–59], which consists of policy evaluation and policy improvement, is introduced as follows. The iterative performance index function $V^{(i)}(x)$ and the iterative control policy $u^{(i)}(x)$ are updated successively with the iteration index i increasing from zero to infinity.

Step 1. Initialization: Let $i = 0$ and start with an initial admissible control policy $u^{(0)}(x)$.

Step 2. Policy evaluation: Solve for $V^{(i+1)}(x)$ by computing the equation as follows:

$$[\nabla V^{(i+1)}(x)]^T(f(x) + g(x)u^{(i)}) + x^TQx + u^{(i)T}Ru^{(i)} = 0. \tag{9}$$

Step 3. Policy improvement: Update the iterative control policy by

$$u^{(i+1)}(x) = -\frac{1}{2}R^{-1}g^T(x)\nabla V^{(i+1)}(x). \tag{10}$$

If $\|V^{(i+1)} - V^{(i)}\| \le \varepsilon$, where ε is a prescribed small positive constant, then stop at Step 3; Else, let $i = i + 1$ and go back to Step 2.

It has been mentioned in [12,14,57–59] that the iterative performance index function $V^{(i)}(x)$ and the iterative control policy $u^{(i)}(x)$ will converge to their optimal values $V^*(x)$ and $u^*(x)$, respectively, as $i \to \infty$.