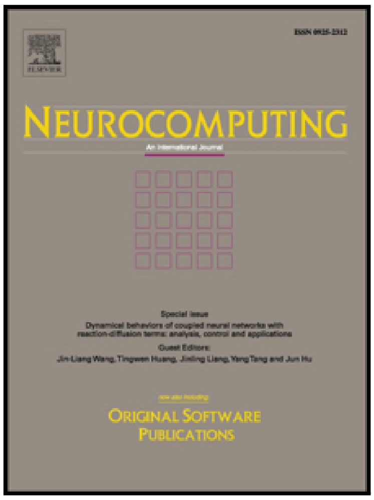

A Data-driven Online ADP Control Method for Nonlinear System Based on Policy Iteration and Nonlinear MIMO Decoupling ADRC

Zhijian Huang , Cheng Zhang , Yanyan Zhang , Huan Zheng , Guichen Zhang

PII: S0925-2312(18)30451-X
DOI: 10.1016/j.neucom.2018.04.024
Reference: NEUCOM 19484

To appear in: *Neurocomputing*

Received date: 17 October 2017
Revised date: 26 February 2018
Accepted date: 25 April 2018

A Data-driven Online ADP Control Method for Nonlinear System Based on Policy Iteration and Nonlinear MIMO Decoupling ADRC

Zhijian Huang[1,3], Cheng Zhang[1], Yanyan Zhang[2], Huan Zheng[1], Guichen Zhang[1]*

(1. Lab of Intelligent Control and Computation, Shanghai Maritime University, Shanghai, 201306, China;

2. The Tenth people's hospital affiliated to Tongji University, Shanghai, 200072, China;

3. Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, Kingston, RI 02881, USA)

Abstract—The action-critic approximate dynamic programming (ADP) depends on its network structure and training algorithm. Since there are some inherent shortcomings of the neural network, this paper proposes a data-driven nonlinear online ADP control method without the neural network. Firstly, a multi-input multi-output (MIMO) policy iteration method is utilized for the proposed ADP. For its policy evaluation, the cost function is approximated with a quadratic function and least square method; for its policy improvement, the optimal control is approximated by solving the quadratic function linearly. In this way, an optimal control equation in form of variable coefficients and system states is deduced. Secondly, a nonlinear MIMO decoupling Active Disturbance Rejection Control method is used to obtain the variable coefficients in real time, which endows the ADP method with a nonlinear performance during its policy improvement. Once the variable coefficients are determined, the data-driven nonlinear ADP control method is deduced. Finally, the examples of an under-actuated nonlinear system and a real application are taken to demonstrate the optimal control effect. Compared with some published methods and their simulation, this method and its simulation excel in the method of policy improvement, nonlinear ability and control performance etc. Thus, the proposed method explores a new way to the ADP, and overcomes the shortcomings of the neural-network-based ADP. Since it enables to work like a PID controller and doesn't require data collecting, training or extra learning, this proposed ADP is a real data-driven nonlinear online optimal control method.

Keywords —data driven, online, approximate dynamic programming, linear quadratic function, least square method, assumed neural network, ADRC.

1. Introduction

The approximate dynamic programming or adaptive dynamic programming (ADP) can obtain the optimal control policy by approximating an associated Hamilton-Jacobi-Bellman (HJB) equation [1]. As an optimal and adaptive control method, it is extensively used to design system controllers [2-3]. The ADP has proved to be effective in many control problems. The optimal control effect of the ADP usually depends on its structure and learning algorithm. There are two main ways to implement the ADP: 1). One way is the neural network (NN)-based ADP; 2). the other way is the non-NN-based ADP. In the past, the action and critic modules of the ADP are mainly approximated with the NN method. However, the NN has some inherent shortcomings although it's suitable to a nonlinear MIMO system. Besides, the non-NN-based ADP way is still very rare. For years, the scholars attempt to explore the ADP that is based on the input and output data without the NN shortcomings, so the data-driven online optimization and the control effect of the ADP can be improved.

As for the NN-based ADP, to solve the poor convergence and parameter shadowing of the gradient descent algorithm of the back propagation (BP) [4], *Govindhasamy et al.* proposed an ADP with the Levenberg-Marquardt algorithm [5]. For a model-free optimization, *Vamvoudakis* proposed a Q-learning algorithm, and a NN-based actor-critic structure is used to approximate its Q-function parameter [6]. *Rajagopal et al.* developed an ADP using the NN to solve a finite horizon stochastic optimal problem. However, it's an offline approach [7]. For a robust performance, *Yang et al.* presented a novel robust ADP regulation method for a class of continuous-time nonlinear systems subject to unmatched perturbations [8], and *Wei et al.* gave a novel convergence analysis for the discrete-time deterministic Q-learning [9]. The applications of the NN-base ADP can be seen in refs.[10-12] etc.

As for the data-driven ADP, after realizing the disadvantage of the NN-based ADP during comparation, *Lee et al.* proposed a data-driven ADP algorithm based on the *k*-Nearest Neighbor method [13]. Though this method is better in its iterative convergence effect, it needs a large calculation. Thus, *Kiumarsi et al.* developed an iterative ADP to on-line solve the linear quadratic (LQ) tracking