Applied Soft Computing 11 (2011) 2946–2963

Contents lists available at ScienceDirect

Applied Soft Computing

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

Two coupled neural-networks-based solution of the Hamilton–Jacobi–Bellman equation

Najla Krichen Masmoudi*, Chokri Rekik, Mohamed Djemel, Nabil Derbel

Research Unit on Intelligent Control, Design and Optimisation of Complex Systems (ICOS), University of Sfax, ENIS, B.P 1173 - 3038, Sfax, Tunisia

ARTICLE INFO

Article history:
Received 31 March 2008
Received in revised form
19 September 2010
Accepted 28 November 2010
Available online 14 December 2010

Keywords:
Hamilton–Jacobi–Bellman (HJB) equation
Neural networks (NN)
Nonlinear discrete time systems
Learning algorithm

ABSTRACT

This work is aimed at looking into the determination of optimal neuro-feedback control for discrete time nonlinear systems. The basic idea consists in the use of two coupled neural networks to approximate the solution of the Hamilton–Jacobi–Bellman equation (HJB) and to obtain a robust feedback closed-loop control law. The used learning algorithm is a modified version of the backpropagation one. As an illustration, a numerical nonlinear discrete time example is considered. Simulation results show the effectiveness of the proposed method.

1. Introduction

In the optimal control theory, the optimal solution of a nonlinear control problem may be computed by solving the Hamilton–Jacobi–Bellman equation (HJB), which is, in general, a difficult task[6,21,35,34].

In the literature, there are many techniques which are interested in the approximation of HJB solution. Kleinman [18] proposed an iterative method to solve Riccati equation for nonlinear systems by successively solving a sequence of Lyapunov equations. Saridis and Lee [33] extended this iterative approach to the case of nonlinear systems. They obtained the optimal control of continuous nonlinear systems by successively solving the generalized Hamilton–Jacobi–Bellman (GHJB) equation. Beard and Saridis [2] used Galerkin's spectral method to approximate a solution to the GHJB equation at each iteration in continuous time. Lyshevski [24] proposed to used a generalized non quadratic functional to give bounded control structure to solve the HJB equation. However, it is difficult to solve the final linear HJB equation.

Sundar and Shiller [36] used the HJB theory to solve the on-line obstacle avoidance problem. The optimal trajectory is generated by following the negative gradient of the return function which is the solution of the HJB equation. For multiple obstacles, avoiding obstacles, optimally one at a time, is equivalent to follow the pseudoreturn function, which is an approximation of the return function for the multi-obstacle problem.

In recent years, there has been a growing research using neural networks. Many researches have been done to exploit them for the control of nonlinear systems [14]. Neural networks have a massively parallel distributed structure, an ability to learn and generalize and a built-in capability to adapt their synaptic weights to changes in surrounding environments [13,40,29]. Neural networks are inherently nonlinear, having very important properties, particularly if the systems underlying physical mechanisms are highly nonlinear. More importantly, they can approximate any nonlinear function with any desirable accuracy [34,19,28,15].

In this context, Miller et al. [25] proposed to used neural network to compute optimal control laws using the HJB equation. Kim et al. [17] proposed explicit solutions to the HJB equation by solving an algebric Riccati equation. They used neural networks to estimate nonlinear uncertainties, adaptively. The neural network is expected to improve performances face to unknown nonlinearities by adding nonlinear effects to the linear optimal controller. Munos et al. [26] combined the gradient descent approach with the HJB equation from conventional optimal control. They used a neural network to approximate the solution of the Hamilton–Jacobi–Bellman (HJB) equation. They derived

* Corresponding author. Tel.: +216 74 274 088; fax: +216 74 275 595.
E-mail addresses: najla_krichen@yahoo.fr (N. Krichen Masmoudi), chokri.rekik@enis.rnu.tn (C. Rekik), mohamed.Djemel@enis.rnu.tn (M. Djemel), nabil.derbel@ieee.org (N. Derbel).

1568-4946/$ – see front matter
doi:10.1016/j.asoc.2010.11.015

the gradient descent rule to integrate this equation inside the domain, for given boundary conditions. Abu-Khalaf and Lewis [1] proposed an approach based on successive approximation techniques in the least-squares in order to solve the GHJB equation, by employing a nonquadratic functional. A neural network is used to identify the GHJB solution. Lewis and Abu-Khalaf [22] employed nonquadratic performance functional to solve constrained control problems for general affine nonlinear continuous-time systems. They formulated the Hamilton-Jacobi-Bellman (HJB) equation using nonquadratic functionals then they solved it using successive iterations between the control and the cost function. Neural networks are used to approximate the solution for the cost function at each successive iteration. Cheng and Lewis [9] proposed fixed-final time constrained input optimal control laws for general affine in the input nonlinear systems by using neural networks to solve the Hamilton–Jacobi–Bellman (HJB) equation. This neural network is constructed to approximate the time-varying cost function by applying the least-squares method on a pre-defined region. It solved approximately the time-varying HJB equation for constrained control nonlinear systems. Ferrari and Stengel [10] proposed a method based on adaptive critic designs to solve the HJB equation in discrete-time. The cost function and control are update using heuristic dynamic programming. Chen and Jagannathan [7,8] used neural networks to obtain nearly optimal solutions to the control of nonlinear discrete-time systems. The approach is based on the least squares successive method to approximate the solution of the generalized Hamilton–Jacobi–Bellman (GHJB) equation in discrete-time on a well-defined attraction region. Successive approximations using the GHJB have not been applied for nonlinear discrete-time systems. A neural network is used to identify the GHJB solution. The result is a closed-loop control based on a neural network that has been tuned a priori in off line mode.

Optimal control problems and its solution using value functions and dynamic programming is named adaptive critic learning. Prokhorov and Wunsch [31] discussed a variety of Adaptive Critic Designs (ACDs) for neurocontrol. These are considered as generalizations of dynamic programming for neural reinforcement learning approaches. A family of Adaptive Critic Designs was proposed by Werbos [39]. These are considered as an optimization technique which combine together concepts of reinforcement learning and approximate dynamic programming. The goal of each design is to approximate the cost-to-go function of the Bellman equation of dynamic programming or some function related to it, and then to find the optimal solution by applying a reinforcement learning algorithm. There are three basic implementations of ACDs called Heuristic Dynamic Programming (HDP), Dual Heuristic Programming (DHP), and Globalized Dual Heuristic Programming [31]. Sutton and Barto [37] cited the main elements to define a reinforcement learning (RL) system: a policy, a reward function, a value function, and, optionally, a model of the environment. The reinforcement learning adopts three threads: learning by 'trial and error' mechanism, problems of optimal control, and temporal-difference methods. In ACDs architecture, we have two neural networks called critic and action which are connected together directly or through an identification model of a plant. The action network minimize an approximation of the cost-to-go. The critic network is trained to estimate the cost-to-go function from the Bellman equation of dynamic programming or some function related to it. Actual and desired outputs of the plant serve as inputs for the Critic network. The aim is to train the action network to minimize this estimate over time, thereby computing an optimal control law [38].

Adaptive critic controllers provide practical approach to attain optimality in the most general case. Optimal control law and value function are modelled as parametric structures (e.g., computational neural networks). By solving the recurrence relation of dynamic programming, there shapes are improved over time. The parametric structures change system parameters without explicit parameter identification. This requires an approximate model of the plant dynamics that admits satisfactory estimates of the future cost [10].

In this paper, we will consider the optimal control problem of nonlinear discrete time systems. We will provide from performance benefits from the HJB equation (e.g., optimality and feedback) and benefits from the numerical properties of neural networks. So, two coupled neural networks will be used to approximate the solution of the discrete HJB equation and the optimal control law. In other words, neural networks will be trained to find the function value, solution of the HJB, and the optimal feedback control law. Weights of the two coupled neural networks are updated using a modified version of the, well known, backpropagation algorithm [30]. The adjustment of weights, based on the gradient descent method, depends on the differentiation of the criterion, to be minimized, with respect to the weights. This differential is also function of the outputs of the two coupled neural networks. Simulation results show that the proposed method gives satisfactory performances. The robustness of the proposed approach is verified.

Five sections comprise this paper. Section 2 is an overview of the optimal control theory. Section 3 presents the proposed approach. In Section 4, a nonlinear discrete second order numerical example is considered to prove the effectiveness of the proposed method.

2. Optimal control theory

A discrete time nonlinear system can be described by the following state equation:

$$x_{k+1} = f(x_k, u_k) \tag{1}$$

where $x_k \in \mathbb{R}^n$ represents the state vector, and $u_k \in \mathbb{R}^{n_s}$ is the control vector.

The problem is to find the optimal feedback control that minimizes the following criterion:

$$J = \sum_{0}^{\infty} r(x_k, u_k) \tag{2}$$

To do numerically, we consider the following modification on the criterion:

$$J = S(x_k, N) + \sum_{k=0}^{N-1} r_k(x_k, u_k) \tag{3}$$

with a finite time horizon N. The function $S(.)$ penalizes the state x_N at the time horizon $k = N$, and the function $r(.)$ penalizes the states x_k and the control inputs u_k through $k = 0$ to $k = N - 1$.