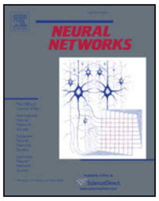
Neural Networks 22 (2009) 237–246

Contents lists available at ScienceDirect

Neural Networks

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



2009 Special Issue

Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems

Draguna Vrabie*, Frank Lewis

Automation and Robotics Research Institute, University of Texas at Arlington, 7300 Jack Newell Blvd. S., Fort Worth, TX 76118, USA

ARTICLE INFO

Article history:
Received 11 January 2009
Received in revised form 16 March 2009
Accepted 19 March 2009

Keywords:
Direct adaptive optimal control
Policy iteration
Neural networks
Online control

ABSTRACT

In this paper we present in a continuous-time framework an online approach to direct adaptive optimal control with infinite horizon cost for nonlinear systems. The algorithm converges online to the optimal control solution without knowledge of the internal system dynamics. Closed-loop dynamic stability is guaranteed throughout. The algorithm is based on a reinforcement learning scheme, namely Policy Iterations, and makes use of neural networks, in an Actor/Critic structure, to parametrically represent the control policy and the performance of the control system. The two neural networks are trained to express the optimal controller and optimal cost function which describes the infinite horizon control performance. Convergence of the algorithm is proven under the realistic assumption that the two neural networks do not provide perfect representations for the nonlinear control and cost functions. The result is a hybrid control structure which involves a continuous-time controller and a supervisory adaptation structure which operates based on data sampled from the plant and from the continuous-time performance dynamics. Such control structure is unlike any standard form of controllers previously seen in the literature. Simulation results, obtained considering two second-order nonlinear systems, are provided.

1. Introduction

In an environment in which a number of players compete for a limited resource, optimal behavior with respect to desired long term goals leads to long term advantages. In a control engineering framework the role of the environment is played by a system to be controlled (this ranges from industrial processes such as distillation columns and power systems, to airplanes, medical equipment and mobile robots), while the controller, equipped with sensors and actuators, plays the role of the agent which is able to regulate the state of the environment such that desired performances are obtained. An intelligent controller is able to adapt its actions in front of unforeseen changes in the system dynamics. In the case in which the controller has a fixed parametric structure, the adaptation of controller behavior is equivalent to changing the values of the controller parameters. From a control engineering perspective, not every automatic control loop needs to be designed to exhibit intelligent behavior. In fact in industrial process control there exists a hierarchy of control loops which has at the lowest level the simplest and most robust regulation, which provides fast reaction in front of parametric and non-parametric disturbances without controller adaptation, while at the topmost end are placed the so-called money-making loops, whose operation close to optimality has the greatest impact on maximization of income. In the latter case the control performance is not explicitly defined in terms of desired trajectories for the states and/or outputs of the system; instead it is implicitly expressed through a functional that captures the nature of the desired performance in a more general sense. Such an optimality criterion characterizes the system's performance in terms of the control inputs and system states; it is in fact an implicit representation of a desired balance between the amount of effort invested in the control process and the resulting outputs.

Optimal control refers to a class of methods that can be used to synthesize a control policy which results in best possible behavior with respect to the prescribed criterion (i.e. control policy which leads to maximization of performance). The solutions of optimal control problems can be obtained either by using Pontryagin's minimum principle, which provides a necessary condition for optimality, or by solving the Hamilton–Jacobi–Bellman (HJB) equation, which is a sufficient condition (see e.g. Kirk (2004) and Lewis and Syrmos (1995)). Although mathematically elegant, both approaches present a major disadvantage posed by the requirement of complete knowledge of the system dynamics. In the case when only an approximate model of the system is available, the optimal controller derived with respect to the system's model will not perform optimally when applied for the control of the real process. Thus, adaptation of the controller

* Corresponding author. Tel.: +1 817 272 5938; fax: +1 817 272 5938.
E-mail addresses: dvrabie@uta.edu (D. Vrabie), lewis@uta.edu (F. Lewis).

0893-6080/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2009.03.008

parameters such that operation becomes optimal with respect to the behavior of the real plant is highly desired.

Adaptive optimal controllers have been developed either by adding optimality features to an adaptive controller (e.g. the adaptation of the controller parameters is driven by desired performance improvement reflected by an optimality criterion functional) or by adding adaptive features to an optimal controller (e.g. the optimal control policy is improved relative to the adaptation of the parameters of the model of the system). A third approach to adaptive optimal control (Sutton, Barto, & Williams, 1992), namely reinforcement learning (RL) (Sutton & Barto, 1998), was introduced and extensively developed in the computational intelligence and machine learning societies, generally to find optimal control policies for Markovian systems with discrete state and action spaces (Howard, 1960). RL-based solutions to the continuous-time optimal control problem have been given in Baird (1994) and Doya (2000). The RL algorithms are constructed on the idea that successful control decisions should be remembered, by means of a reinforcement signal, such that they become more likely to be used a second time. Although the idea originates from experimental animal learning, where it has been observed that the dopamine neurotransmitter acts as a reinforcement informational signal which favors learning at the level of the neuron (see e.g. Doya, Kimura, and Kawato (2001) and Schultz, Tremblay, and Hollerman (2000)), RL is strongly connected from a theoretical point of view with direct and indirect adaptive optimal control methods. In the present issue, Werbos (2009) reviews four generations of general-purpose learning designs in Adaptive, Approximate Dynamic Programming, which provide approximate solutions to optimal control problems and include reinforcement learning as a special case. He argues the relevance of such methods not only for the general goal of replicating human intelligence but also for bringing a solution of efficient regulation in electrical power systems.

The main advantage of using RL for solving optimal control problems comes from the fact that a number of RL algorithms, e.g. Q-learning (Watkins, 1989) (also known as Action Dependent Heuristic Dynamic Programming (Werbos, 1989, 1992)), do not require knowledge or identification/learning of the system dynamics. This is important since it is well known that modeling and identification procedures for the dynamics of a given nonlinear system are most often time consuming iterative approaches which require model design, parameter identification and model validation at each step of the iteration. The identification procedure is even more difficult when the system has hidden nonlinear dynamics which manifest only in certain operating regions. In the RL algorithms' case the learning process is moved at a higher level having no longer as an object of interest the system's dynamics but a performance index which quantifies how close to optimality the closed-loop control system operates. In other words, instead of identifying a model of the plant dynamics, that will later be used for the controller design, the RL algorithms require identification of the static map which describes the system performance associated with a given control policy. One sees now that, as long as enough information is available to describe the performance associated with a given control policy at all significant operating points of the control system, the system performance map can be easily learned, conditioned by the fact that the control system maintains stability properties. This is again advantageous compared with an open-loop identification procedure which, due to the excitatory inputs required for making the natural modes of the system visible in the measured system states, could have as a result the instability of the system.

Even in the case when complete knowledge of the system dynamics is available, a second difficulty appears from the fact that the HJB equation, underlying the optimal control problem, is generally nonlinear and most often does not possess an analytical solution; thus the optimal control solution is regularly addressed by numerical methods (Huang & Lin, 1995). Also from this point of view, RL algorithms provide a natural approach to solve the optimal control problem, as they can be implemented my means of function approximation structures, such as neural networks, which can be trained to learn the solution of the HJB equation. RL algorithms, such as the one that we will present in Section 3, and develop for online implementation in Section 4, can easily be incorporated in higher-level decision making structures of the sort presented in (Brannon, Seiffertt, Draelos, & Wunch, 2009).

RL algorithms can be implemented on Actor/Critic structures which involve two function approximators, namely the Actor, which parameterizes the control policy, and the Critic, a parametric representation for the cost function which describes the performance of the control system. In this case the solution of the optimal control problem will be provided in the form of the Actor neural network for which the associated cost, i.e. the output of the Critic neural network, has an extremal value.

In this paper we present an adaptive method, which uses neural-network-type structures in an Actor/Critic configuration, for solving online the optimal control problem for the case of nonlinear systems, in a continuous-time framework, without making use of explicit knowledge on the internal dynamics of the nonlinear system. The method is based on Policy Iteration (PI), an RL algorithm which iterates between the steps of policy evaluation and policy improvement. The PI method starts by evaluating the cost of a given admissible initial policy and then uses this information to obtain a new control policy, which is improved in the sense of having a smaller associated cost compared with the previous policy, over the domain of interest in the state space. The two steps are repeated until the policy improvement step no longer changes the present policy, this indicating that the optimal control behavior is obtained.

In the case of continuous-time systems with linear dynamics, PI was employed for finding the solution of the state feedback optimal control problem (i.e. LQR) in Murray, Cox, Lendaris, and Saeks (2002), while the convergence guarantee to the LQR solution was given in Kleinman (1968). The PI algorithm, as used by Kleinman (1968), requires repetitive solution of Lyapunov equations, which involve complete knowledge of the system dynamics (i.e. both the input-to-state and internal system dynamics specified by the plant input and system matrices). For nonlinear systems, the PI algorithm was first developed by Leake and Liu (1967). Three decades later it was introduced in Beard, Saridis, and Wen (1997) as a feasible adaptive solution to the CT optimal control problem. In Beard et al. (1997) the Generalized HJB equations (a sort of nonlinear Lyapunov equations), which appear in the PI algorithm, were solved using successive Galerkin approximation algorithms. A neural-network-based approach was developed and extended to the cases of H2 and H-infinity with constrained control in Abu-Khalaf and Lewis (2005) and Abu-Khalaf, Lewis, and Huang (2006). Neural-network-based Actor/Critic structures, in a continuous-time framework, with neural network tuning laws have been given in Hanselmann, Noakes, and Zaknich (2007). All of the above-mentioned methods require complete knowledge of the system dynamics.

In Vrabie, Pastravanu, and Lewis (2007) and Vrabie and Lewis (2008) the authors gave a new formulation of the PI algorithm for linear and nonlinear continuous-time systems. This new formulation allows online adaptation (i.e. learning) of the continuous-time operating controller to the optimal state feedback control policy, without requiring knowledge of the system internal dynamics (knowledge regarding the input-to-state dynamics is still required, but from a system identification point of view this knowledge is relatively easier to obtain).