Learning to Control an Inverted Pendulum Using Neural Networks

Charles W. Anderson

ABSTRACT: An inverted pendulum is simulated as a control task with the goal of learning to balance the pendulum with no a priori knowledge of the dynamics. In contrast to other applications of neural networks to the inverted pendulum task, performance feedback is assumed to be unavailable on each step, appearing only as a failure signal when the pendulum falls or reaches the bounds of a horizontal track. To solve this task, the controller must deal with issues of delayed performance evaluation, learning under uncertainty, and the learning of nonlinear functions. *Reinforcement* and *temporal-difference* learning methods are presented that deal with these issues in order to avoid unstable conditions and balance the pendulum.

Introduction

The inverted pendulum is a classic example of an inherently unstable system. Its dynamics are basic to tasks involving the maintenance of balance, such as walking and the control of rocket thrusters. A number of control design techniques have been investigated using the inverted pendulum [1]-[4]. The successful application of these design techniques requires considerable knowledge of the system to be controlled, including an accurate model of the dynamics of the system and an expression of the system's desired behavior, usually in the form of an objective function.

How can control be accomplished when such knowledge is not available? This question is addressed here by considering the inverted pendulum control problem when the dynamics are not known a priori and an analytical objective function is not given. All that is known are the values and ranges of the state variables of the inverted pendulum system and that a negative *failure signal* is to be maximized over time. A function that selects control actions given the current state of the pendulum must be learned through experience by trying various actions and noting the results, starting with no hints as to which actions are correct.

Without an objective function to evaluate states and actions, modifications to the controller can be based only on the occurrence of failure signals. A long sequence of actions can develop before a failure signal is encountered, resulting in the difficult *assignment-of-credit* problem, where it is necessary to decide which actions in the sequence contributed to the failure.

In this paper, neural network learning methods are described that learn to generate successful action sequences by acquiring two functions: an *action function*, which maps the current state into control actions, and an *evaluation function*, which maps the current state into an evaluation of that state. The evaluation function is used to assign credit to individual actions. Two networks having a similar structure are used to learn the action and evaluation functions. They will be referred to as the *action network* and the *evaluation network*.

As shown in later sections, the desired evaluation function for the inverted pendulum task is nonlinear; a single-layer neural network cannot form this map. One solution to this problem is to transform the original state variables into a new representation with which a single-layer network can form the evaluation function. Barto et al. [5] demonstrated a quantization of the state space of the inverted pendulum with which single-layer networks could learn to balance the pendulum. A second solution is to add a second adaptive layer that learns such a representation. Anderson [6] extended the work of Barto et al. by applying a form of the popular error back-propagation method to two-layered networks that learn to balance the pendulum given the actual state variables of the inverted pendulum as input.

In this paper, the work of Barto et al. and Anderson is summarized by discussing the neural network structures and learning methods from a functional viewpoint and by presenting the experimental results. First, the inverted pendulum task and previous applications of neural networks to this task are described.

Presented at the 1988 American Control Conference, Atlanta, Georgia, June 15–17, 1988. Charles W. Anderson is with the Self-Improving Systems Department of GTE Laboratories, Inc., Waltham, MA 02254.

Inverted Pendulum

The inverted pendulum task involves a pendulum hinged to the top of a wheeled cart that travels along a track, as shown in Fig. 1. The cart and pendulum are constrained to move within the vertical plane. The state at time t is specified by four real-valued variables: the angle between the pendulum and vertical and the angular velocity (θ_t and $\dot{\theta}_t$) and the horizontal position and velocity of the cart (h_t and $\dot{h}_t$). The inverted pendulum system was simulated using the following equations of motion, where the units of θ, h, and time t are radians, meters, and seconds, respectively, and where g is the acceleration due to gravity (9.8 m/sec^2), F_t the output of the action network (± 10 N), m_c the mass of the cart (1.0 kg), m the mass of the pendulum plus the cart (1.1 kg), and l the distance from the pivot to the pendulum's center of mass (0.5 m).

$$\ddot{\theta}_t = \frac{mg \sin \theta_t - \cos \theta_t [F_t + m_p l \dot{\theta}_t^2 \sin \theta_t]}{(4/3)ml - m_p l \cos^2 \theta_t}$$

$$\ddot{h}_t = \{F_t + m_p l [\dot{\theta}_t^2 \sin \theta_t - \ddot{\theta}_t \cos \theta_t]\}/m$$

This system was simulated by numerically approximating the equations of motion using Euler's method with a time step of $\tau = 0.02$ sec and discrete-time state equations of the form $\theta[t+1] = \theta[t] + \tau\dot{\theta}[t]$. The sampling rate of the inverted pendulum's state and the rate at which control forces are applied are the same as the basic simulation rate, i.e., 50 Hz.

The goal of the inverted pendulum task is to apply a sequence of right and left forces of fixed magnitude to the cart such that the pendulum is balanced and the cart does not hit the edge of the track. A zero-magnitude

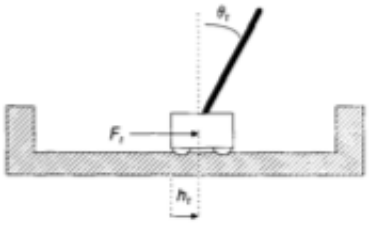

Fig. 1. The inverted pendulum.