J. King Saud Univ., Vol. 10, *Comp. & Info. Sci.*, pp. 73-80 (A.H. 1418/1998)

Truck Backer-Upper Control Using Adaptive Critic Learning

M. Khorsheed[*] and M. A. Al-Sulaiman[**]

[*] *Ministry of Defense and Aviation, P.O. 6917, Riyadh 11452, Saudi Arabia*

[**] *Computer Engineering Dept., College of Computer & Information Sciences, King Saud University, P.O. Box 51178, Riyadh 11543, Saudi Arabia*

(Received 08 April 1996, accepted for publication 01 October 1996)

Abstract. Backing up a truck to a loading dock is a difficult nonlinear control problem. Interesting applications of backpropagation (BP) learning to solve this problem have been proposed. However, BP requires extensive training. In this paper, we propose a solution based on the adaptive critic learning. The proposed scheme is considered as a first attempt towards solving Truck Backer-Upper (TBU) using this kind of learning. Results obtained through computer simulation of TBU are presented.

1. Introduction

For the sake of mathematical tractability, control systems theory has been mainly directed to linear systems or linear approximations of nonlinear systems. However, a linear solution may lead to unexpected performance degradation or instability [1,2]. One such nonlinear control problem is the truck backer-upper (TBU); where a truck is backed to a loading dock starting from an arbitrary position. While a human driver is backing a truck, it is observable that the driver is backing, going forward, backing again, going forward etc., and finally backing to almost the desired position along the dock. The situation becomes more difficult if only backward movements are allowed.

Many neural network-based methods have been proposed to solve the TBU problem [3,4,7]. An interesting application of backpropagation learning was presented by Nguyen and Widrow [4] who trained a neural network to back-up a simulated truck to a loading dock in a planar parking lot. Nguyen and Widrow's solution was based on back-propagation learning through time and required an emulator of the plant. The solution was sensitive to the emulator accuracy and it required the error to be propagated

backward through a sequence of emulator-controller pairs, adjusting controller weights along the way.

In this paper we propose a solution based on the adaptive critic learning [5] as an alternative to back-propagation. The TBU problem is typical of a wide range of control problems, where the outcome of a long sequence of actions is only known at the end [6] (e.g. the truck hits the boundaries) for which the adaptive critic learning has been found eminently suitable.

Figure 1 shows the geometry of TBU problem. There are three state variables x, y, and ϕ. (x,y) specify the position of the rear center of the truck, and ϕ specifies the angle of the truck with the horizontal. Collectively, these will be referred to as the state vector. The state-variable ranges are : $0 \le x, y \le 100 \quad -90 \le \phi \le 270$. At each step the output of the controller will be the steering angle θ, where : $-30 \le \theta \le 30$.

The truck moves backward by some fixed distance at every step, until it hits the loading dock. The goal is to make the truck arrive at the loading dock at position $(x_d,y_d) = (50,100)$ at right angle ($\phi_d = 90$). Simple kinematics equations are used to update the state variables. If the truck moves backward from (x,y,ϕ) to (x',y',ϕ') then :

$$\phi' = \phi + \theta \tag{1}$$
$$x' = x + z*\mathrm{Cos}\phi' \tag{2}$$
$$y' = y + z*\mathrm{Sin}\phi' \tag{3}$$

z denotes the fixed driving distance of the truck for all backing movements.

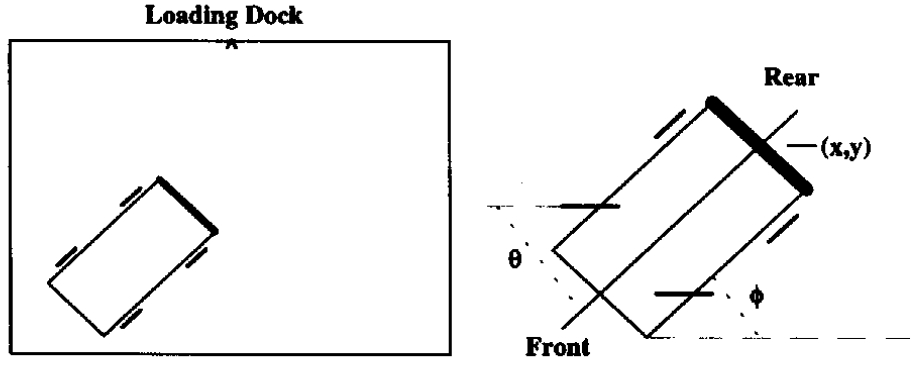

Fig. 1. Schematic diagram of TBU model.

The rest of the paper is organized as follows: In section-2 the adaptive critic system is reviewed. Section-3 introduces two new elements which are added to the controller: a) a reinforcement function to calculate the appropriate punishment signal relevant to the error occurred, and b) a steering angle multiplexer which is a device with a prior knowledge to help in selecting the steering angle's absolute value. Simulation results and concluding remarks are given in section- 4 .

2. Adaptive Critic Learning

Neural learning systems are broadly classified into three types : Supervised learning, Reinforcement learning and Unsupervised learning. Among these systems, reinforcement learning more closely interacts with the environment in real time and learns through favorable and unfavorable environmental reactions. The environmental feedback (the yes/no reinforcement signal) is only evaluative, not instructive.

The adaptive critic system is composed of two neuron-like elements : the adaptive search element (ASE) and the adaptive critic element (ACE). There is a state box decoder which identifies the box that contains the input state vector and distributes the received failure signal r to the controller. The decoder has real-valued input pathways for the system state vector. For the TBU, there are two input pathways x and ϕ. State variables x and ϕ are quantized into a number of regions. There are 11 grades of x-axis truck position, and 9 grades of horizontal angle ϕ. Since the nature of the problem differs from that of the cart-pole problem [5], these grades are not equally divided. For example, as the truck approaches the loading dock, the interval should be smaller than when the truck is far away from the loading dock.

2.1 Associative search element (ASE)

The decoder transforms each state vector into a 99-component binary vector whose components are all zero except for a single one in the position corresponding to the box containing the state vector. This vector, provided as an input to the Associative Search Element (ASE), selects the synapse corresponding to the appropriate box. Associated with each input pathway I is a real-valued weight with value at time t denoted $w_i(t)$. The decoder also distributes to all synapses the information received by the failure signal through reinforcement pathway.

$$u(t) = \begin{cases} 1 & (w_{\text{active-state}}(t) + \text{noise}(t)) \geq 0 \\ 0 & \text{Otherwise} \end{cases} \tag{4}$$

where noise(t) is a real random variable with probability density function of Gaussian distribution and $w_{active\_state}(t)$ is the real-valued associated with active state input pathway. The outputs are determined by the probability biased by the weight of the active state coming from the decoder.

2.2 Adaptive critic element (ACE)

The Adaptive Critic Element (ACE) receives the externally supplied reinforcement signal which it uses to determine how to compute an improved reinforcement signal that it sends to ASE. The ACE's output (the improved or internal reinforcement signal) is computed as follows :

$$\hat{r}(t) = r(t) + \gamma p(t) - p(t-1) \qquad (5)$$

$$0 \leq \gamma \leq 1$$

where p(t) is a prediction of eventual reinforcement at time t.

3. Extended Adaptive Critic Learning

Due to its exceptional promise, the adaptive critic learning has been variously modified and extended. In this section we describe two new elements as extensions to the basic adaptive critic system Fig. 2.

3.1 Reinforcement function

Positive/negative reinforcement indicates the occurrence of a rewarding/punishing event. For the truck backer-upper problem, the environment will keep giving zero reinforcement (r=0) throughout a trial and gives a negative reinforcement signal only when the truck hits any of the boundaries. The punishment signal will not be identical in all cases, i.e. if the truck hits the boundaries near the loading dock then the punishment should be less than if it hits the boundaries away from the loading dock. To achieve this goal, a reinforcement function is used to calculate the appropriate punishment signal relevant to the final position of the truck (x_f, y_f) :

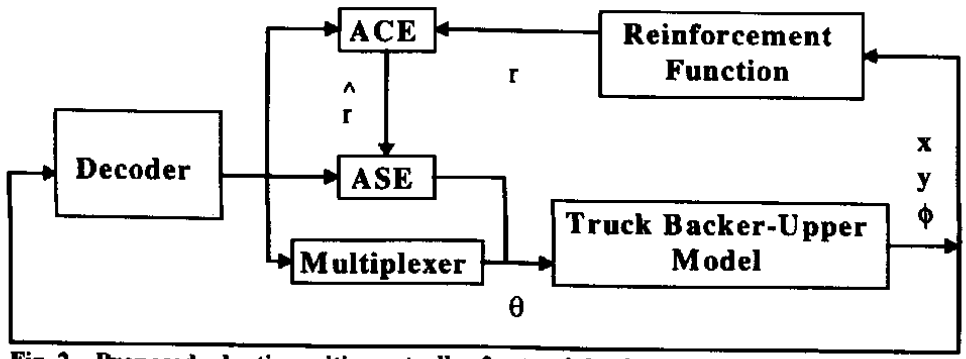

Fig. 2. Proposed adaptive critic controller for truck backer-upper.

$$r = -\sqrt{\left(\frac{x_f}{x_d} - 1\right)^2 + \left(\frac{y_f}{y_d} - 1\right)^2} \tag{6}$$

In Eq.6, the errors of docking the truck in both directions are normalized. This approach accelerates the convergence to the desired position (x_d, y_d).r has the following interpretation : r=0, the truck has reached the loading dock. r<0, the truck's final position is different from the desired one. Graded reinforcement provides an efficient solution to the TBU problem, where different classes of reinforcement signal are needed in order to decide the strength of the punishment signal that should be forwarded.

3.2 Steering-angle multiplexer

The output of the ASE provides the direction (left or right) of the steering angle. Unlike the pole-balancing problem [6,7], the truck backer-upper problem needs to determine the value of the steering angle in addition to the direction. This value is not a constant. It is related to the system state that the truck falls in. So, there should be a way to determine this value. In our proposed adaptive critic method we have included a multiplexer which provides the absolute value of the steering angle. The internal structure of this multiplexer is a simple feedforward neural network consisting of three layers (the input layer, the hidden layer and the output layer). The number of nodes in the input layer is equal to half the number of components output by the decoder. The difference is in the nonzero component, it's value being equal to the product of the membership values of x and ϕ in their intervals which they are falling in. This membership value determines the degree to which the element (x or ϕ) belongs to the interval. It is mathematically represented below: (C_x and C_ϕ represent the intermediate values of the intervals)

$$e_x = \left|\frac{x}{C_x}\right| \tag{7}$$

$$e_\phi = \left|\frac{\phi}{C_\phi}\right| \tag{8}$$

The hidden layer contains 4-nodes. Finally, the output layer (the magnitude value of the steering angle) has only one node.

An important issue regarding the back-propagation learning is to decide about the number of required patterns to train the network. If we divide the planar into four sections, each two sections falling on the same diagonal being equal in magnitude but opposite in sign, then the number of patterns required is less than half the number of possible states. If x and ϕ fall in the intervals x_n and ϕ_n respectively, then the magnitude value will be exactly the same if x and ϕ falls in the intervals x_0 and ϕ_m respectively. This symmetry makes it possible to use only 40 input-output patterns to train the neural network.

4. Results and Discussion

We have presented computer simulation results to verify the application of adaptive critic learning to the truck backer-upper control problem. Some points are clarified before presenting the results :

1. The initial position of the truck is random, and it can also be selected.
2. The fixed driving distance (z) of the truck that mentioned in equations 2 and 3 equals 1.
3. Each run consisted of a sequence of trials, where in each trial the truck begins from the same initial position for that run and stops either with a failure signal or a signal of a success (reaching the loading dock).
4. At the start of each run all the weights are set to zero.
5. At the start of each trial all the trace variables are set to zero.

Figure 3 shows two trajectories of the truck starting from two different initial positions : $(x_0, y_0, \phi_0) = (15,13,157), (73,24,37)$.
The future work will concentrate on extending the proposed scheme to a cab-and-trailer model. This will require more training patterns. Also, the number of input states will increase.

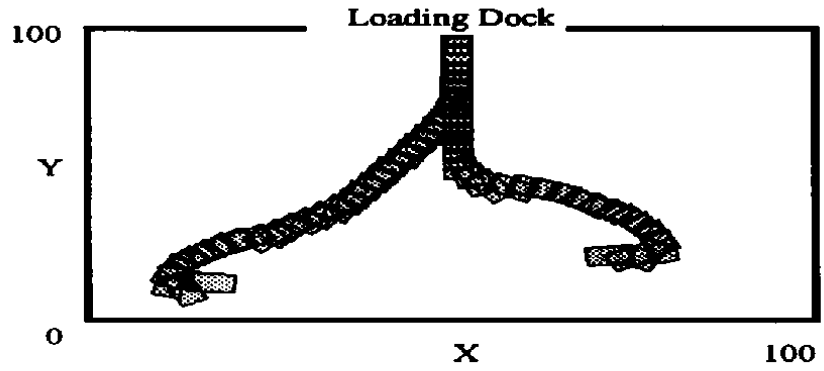

Fig. 3. Trajectories of the truck.

References

[1] Wang, H., Tanaka, K. and Griffin, M. "An Approach to Fuzzy Control of Nonlinear Systems Stability and Design Issues." *IEEE Trans. On Fuzzy Systems*, 4, No. 1 (1996), 14-23.

[2] Levin, A. and Narendra, K. "Control of Nonlinear Dynamical System Using Neural Networks." *IEEE Trans. On Neural Networks*, 7, No. 1 (1996), 30-42.

[3] Jenkins, R. and Yuhas, B. "A Simplified Neural Network Solution Through Problem Decomposition." *IEEE Trans. On Neural Networks*, 4, No. 4 (1993), 718-720.

[4] Nguyen, D. and Widrow, B. "Neural Networks for Self-Learning Control System." *IEEE Control Magazine*, (1990), 18-23.

[5] Barto, A., Sutton, R. and Anderson, C.W. "Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems." *IEEE Trans. on Systems, Man and Cybernetics*, 13 (1983), 834-846.

[6] Ahson, S. and Srinivas, R. "Learning Action Probabilities from Delayed Reinforcement." *Int. Journal Systems SCI.*, 23, No. 2 (1993).

[7] Kosko, B. *Comparison of Fuzzy and Neural Truck Backer-Upper Control System, Neural Networks and Fuzzy Systems.* Englewood Cliffs, NJ : Prentice-Hall, (1992) 339-361.

التحكم بالشاحنة المتراجعة للخلف باستخدام التعلم النقدي التكيّفي

م. خورشيد و م. أ. السليمان

وزارة الدفاع والطيران، ص.ب ٦٩١٧، الرياض ١١٤٥٢، المملكة العربية السعودية

وقسم هندسة الحاسب، كلية علوم الحاسب والمعلومات، جامعة الملك سعود،

ص.ب ٥١١٧٨، الرياض ١١٥٤٣، المملكة العربية السعودية

(قدّم للنشر في ١٩٩٦/٤/٨م؛ وقبل للنشر في ١٩٩٦/١٠/١م)

ملخص البحث . إن تحريك شاحنة إلى الخلف لترجع إلى مرفأ تحميل لهو مسألة تحكّم صعبة وغير خطّية. وقد اقترح استخدام طريقة التعليم بواسطة الانتشار العكسي لحل هذه المسألة. ولكن هذه الطريقة تتطلب إجراء تدريب مكثّف للشبكة العصبية. وفي هذه الورقة نقترح حلاً يعتمد على التعّلم النقدي التكيّفي. وهذا الحل المقترح يعتبر خطوة أوّلية لحل مسألة الشاحنة المتراجعة للخلف باستخدام طرق التعلم هذه. كما تعرض الورقة بعض النتائج التي تمّ التوصل إليها لتطبيق الحل المقترح وذلك عن طريق المحاكاة بواسطة الحاسوب.