Proceedings of the World Congress on Engineering 2008 Vol III
WCE 2008, July 2 - 4, 2008, London, U.K.

Development of a Novel Reinforcement Learning Automata Method for Optimum Design of Proportional Integral Derivative Controller For Nonlinear Systems

S. M. A. Mohammadi [+]***, A. Gharaveisi*** [+]***, M. Mashinchi*** [+] ***and S.M, Vaezi-Nejad*** [*]

Abstract: **This paper is based on a major research project on the development of a novel design of proportional integral derivative (PID) controller for nonlinear systems. The proposed design has superior features, including easy implementation, stable and fast convergence characteristic , and good computational efficiency.**

Key words: CDARLA, DARLA , Optimal control , PID Controller, Reinforced learning.

1. Introduction

The idea that we learn by interacting with our environment is probably the first to occur to us when we think about the nature of learning. Any method that is well suited to solving that problem, we consider to be a reinforcement learning method. Reinforcement learning promises to be an extremely important new technology with immense practical impact and important scientific insights into the organization of intelligent systems [1-3]. Adaptation parameters and tuning the process can be achieved by *Continuous Action Reinforcement Learning Automata* (CARLA) and *Discrete Action Reinforcement Learning Automata* (DARLA) [4]. CARLA was developed as an extension of the discrete stochastic learning automata methodology and it is shown in Fig 1. DARLA replaces the discrete action space with a continuous one, making use of continuous probability distributions and hence making it more appropriate for engineering applications that are inherently discrete.

+ Dept of Electrical Eng, Shahid Bahonar University, Iran.
* Dept of Computing, London Metropolitan University, UK

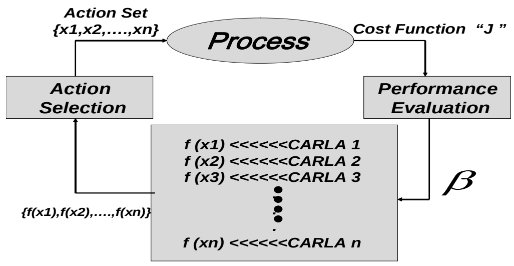

Fig 1: Leaning system by CARLA

The DARLA operates through interaction with a random or unknown environment by selecting actions in astochastic trial and error process. It replaces the discrete action space with a continuous one , making use of continuous probability distributions and hence making it more appropriate for engineering applications that are inherently continuous in nature. Fig 2 shows the learning system by DARLA.

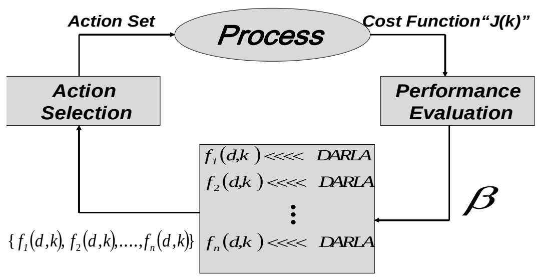

Fig 2: Learning system by DARLA

It should be noted that each CARLA operates on a separate action- typically a parameter value in a model or controller- and the automata set runs in a parallel implementation, determining multiple parameter values. The only interconnection between CARLAs is through the environment and via a shared performance evaluation function. Within each automata, each action has an associated probability density function $f_i(x)$ that is used as the basis for its selection. As described in section 2 , the calculation must be done separately for *'n'* probability density function. The Ball and Beam system belongs to the class of under-actuated mechanical systems having fewer control inputs than degrees of freedom. A ball is placed on a straight beam and rolls back and forth as one end of the beam is raised and lowered by a cam. The position of the ball is controlled by changing the angular position of the cam. This is a second order system, since only the inertia of the ball is taken into account, and not that of the cam or the beam, although the mass of the beam is taken into account in the fourth order state-space model. This renders the control task more challenging making the Ball and Beam system a classical benchmark for testing different control techniques. Fig 3 and table 1 show figure, variables and parameters of system, respectively. In this paper, besides demonstrating how to employ the proposed reinforcement learning method to obtain the optimal PID controller parameters of the system, it will be shown that the proposed method has a better performance than the conventional methods in solving the optimal PID controller parameters.

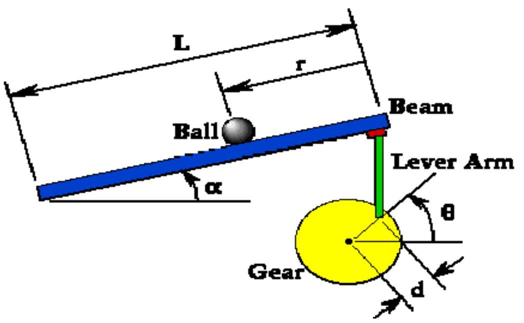

Fig 3: Ball and Beam system

Table 1 shows variables and parameters of the system.

Table 1 : Parameters of Ball&Beam system

M	mass of the ball
R	radius of the ball
d	lever arm offset
g	gravitational acceleration
L	length of the beam
J	ball's moment of inertia
r	ball position coordinate
alpha	beam angle coordinate
theta	servo gear angle

2. Outline of the Proposed Reinforcement Learning Method

Let ' n ' be the number of parameters which must be adjusted so that the index function approaches to minimum value . We can consider each state of the environment (system) in the ' n ' dimensional space and use the common discrete probability function (CDPF) $f_{X1,X2,\ldots\ldots,Xn}(x_1,x_2,\ldots\ldots,x_n)$ for each cell. This must be stored and updated to a new value or zero at discrete sample points. A typical layout for the proposed method is shown in Fig 4.

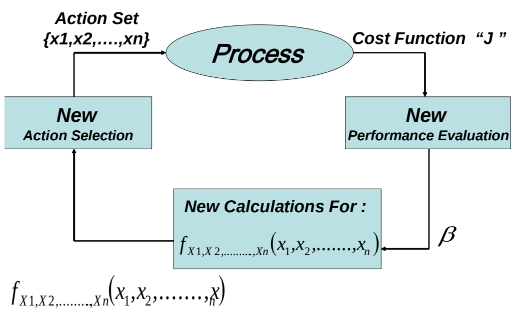

Fig 4 Learning system by Proposed Algorithm

In this method, instead of calculating $f_i(d)$ ' n ' times, we calculate one matrix function so that the speed of convergence will increase .Let n=3 and consider a 3–dimensional probability functional cubic space that complies with above definition. The number of distinct values at each

ISBN:978-988-17012-4-4

dimension is denoted as *Ndiv1*,*Ndiv2* and *Ndiv3*, respectively. By using this structure, task – relevant data and probability functional value of each cell can be precomputed and materialised into desired value which will be demonstrated. A 3– dimensional data cube that complies with above definition is shown in Fig. 5 .

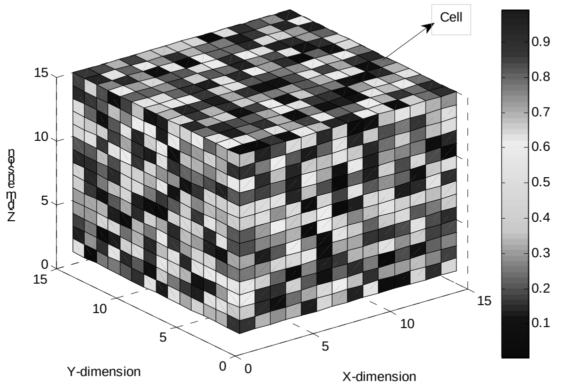

Fig. 5 A random ' *3'* dimensional probability functional cube .

The $f_{X1,X2,\dots\dots\dots,Xn}(x_1, x_2, \dots\dots, x_n)$ has more information about the interaction among the variables than the $f_X(x)$. Within each automata, each action has an associated probability density function $f_X(x)$ that is used as the basis for its selection. Action sets, that improve system performance, invoke a high-performance score ; thus, their probability of reselection increases through the learning sub-system. This is achieved by modifying $f_X(x)$ through the use of a Gaussian neighborhood function centered on the successful action. The neighborhood function increases both the probability of the original action, and the probability of actions close to the one selected, where it is assumed that the performance level over a range in each action is continuous and slowly varying. As the system learns, the probability distribution generally converges to a single Gaussian distribution around the desired parameter value. With all *n* actions selected, the set is evaluated in the environment for a suitable time, and a scalar cost value *J*(*k*) is calculated. The calculation is in accordance with some predefined cost function where the cost *J*(*k*) is compared with a memory set of previous minimum values. After defining performance evaluation, each probability density function is updated according to a specified rule. The following flowchart shows how does the method works .

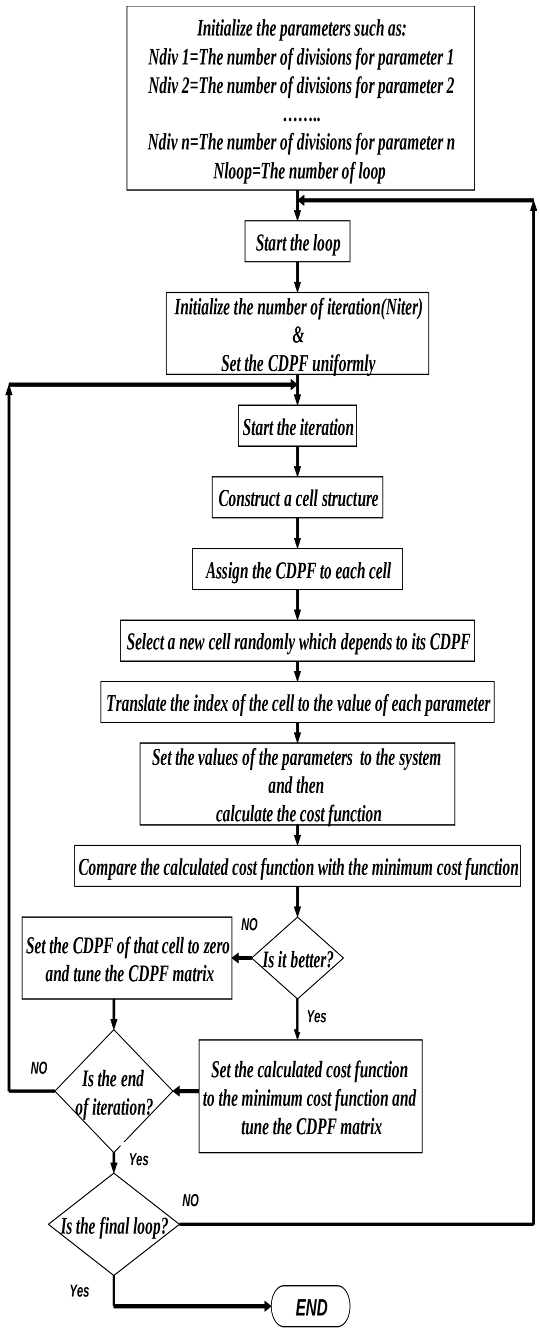

Fig. 6 The flowchart of the proposed reinforcement learning method

The proposed action reinforcement learning automata has been successfully applied to determine PID parameters for Ball and Beam system (idle-speed control), both in simulation and in practice. The method does not require a priori knowledge of the system dynamics and it provides optimized control of complex nonlinear systems.

3. The Proposed Reinforcement Learning PID controlled for nonlinear Systems

The open-loop transfer function of the plant for the ball and beam experiment is given below:

$$\frac{R(s)}{\Theta(s)} = -\frac{mgd}{L(\frac{J}{R^2}+m)} * \frac{1}{s^2}$$

The design criteria for this problem are:

- Settling time less than 7 seconds
- Overshoot less than 5%

The block diagram for this example with a controller and unity feedback of the ball's position is shown below:

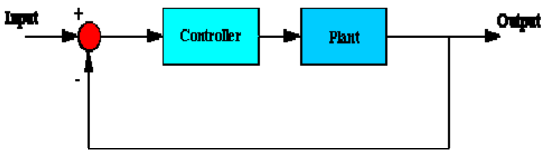

Fig.7.The block diagram of the controller and plant

Recall, that the transfer function for a PID controller is:

$$K_P + \frac{K_I}{s} + K_D s = \frac{K_D s^2 + K_P s + K_I}{s}$$

Dimensional analysis and numerical optimization methods were used to simplify the procedure of obtaining optimal relations. The algorithm proposed here has a clear advantage to the conventional DARLA method of tuning PID controllers [6]. An important and interesting approach, originated by Ziegler and Nichols (1942) and extended since 1984 by Astrdm and Hagglund [7], calls for the tuning of the PI and PID controllers from the identification of a point on the frequency characteristics of the plant (gain and phase). In addition, robustness studies proved the robustness of our new method in comparison with other methods. For mathematical modeling and transfer function of the four components, these components must be linearized, which takes into account the major time constant and ignores the saturation or other nonlinearities. The transfer function of these components can be easily derived. In general, the PID controller design method uses the integrated absolute error (IAE), or the integral of squared-error (ISE), or the integrated of time-weighted-squared-error (ITSE). It is often employed in control system design because it can be evaluated analytically in the frequency domain.

The three integral performance criteria in the frequency domain have their own advantage and disadvantages. For example a disadvantage of the IAE and ISE criteria is that its minimization can result in a response with relatively small overshoot but a long settling time because the ISE performance criterion weights all errors equally independent of time. ITSE performance criterion can overcome the disadvantage of the ISE involving complex and time-consuming deviation of the analytical formula. A novel performance criterion *J(K)* in the time domain is proposed for evaluating the PID controller as:

$$\min_k J(K) = G_e \int_0^T e^2(t)dt + G_u \int_0^T u_c^2(t)dt$$
$$+ G_M M_P + G_s E_{ss} + G_d \sup \left| \frac{de(t)}{d(t)} \right|$$

Where T is the total simulation time and its about $T=20s$, $e(t)$ is tracking error , $u_c(t)$ is control input , M_p is overshoot , E_{ss} is steady-state error and G coefficients are the weight elements.

The two proposed controllers and their performance evaluation criteria in the time domain were implemented by Matlab an control system toolbox, and executed on a *Pentium IV 2.8GHZ personal computer with 256 – MB RAM.*

4 . Simulation

The simulation results of a ball and beam system is beyond the scope of this paper but they clearly show that the proposed controller can perform an efficient search for the optimal PID controller parameters. Therefore, the proposed method has more robust stability and efficiency and can solve the searching and tuning problems of PID controller parameters more easily and quickly than the other methods such as conventional DARLA and CARLA and Ziegler and Nichols. The simulation results also show that the convergence speed of the proposed method is better than conventional methods.

Table 2 : Parameters of Ball&Beam system for simulation

M	mass of the ball	0.11 kg
R	radius of the ball	0.015 m
d	lever arm offset	0.03 m
g	gravitational acceleration	9.8 m/s^2
L	length of the beam	1.0 m
J	ball's moment of inertia	9.99e-6 kgm^2
r	ball position coordinate	
alpha	beam angle coordinate	
theta	servo gear angle	

The following graph shows the system's response to a unit step function .

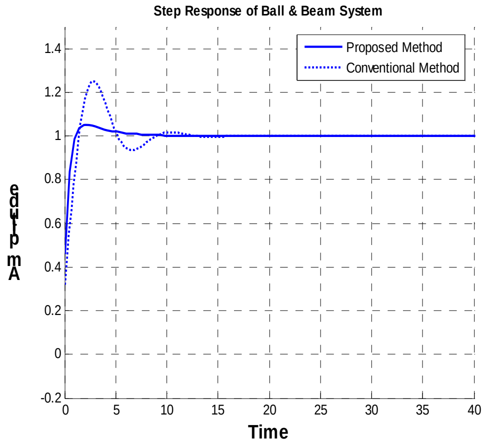

Fig.8. The step response of the system for proposed and conventional method

It is shown that that the settling time and overshoot in the proposed method is better than the conventional DARLA. Let set the tunned parameters as Kp=20.85, Ki=18.95, Kd=3.4 by the proposed method .The resulting graphs for the performance index in the first, second and third loops are as shown in the following figures.

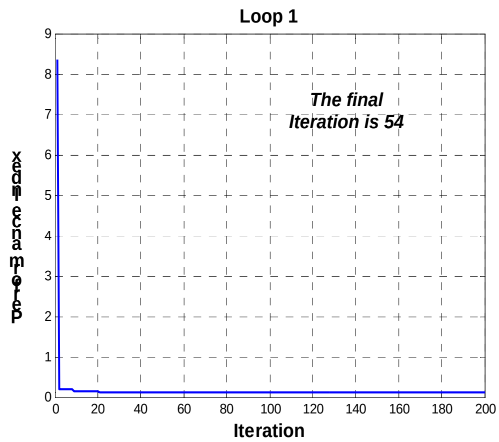

Fig.9. The performance index in the first loop

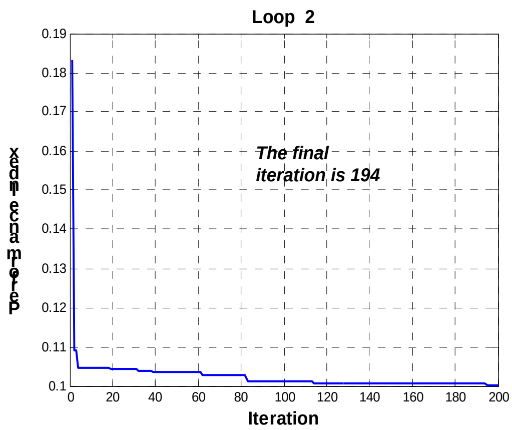

Fig.10. The performance index in the second loop

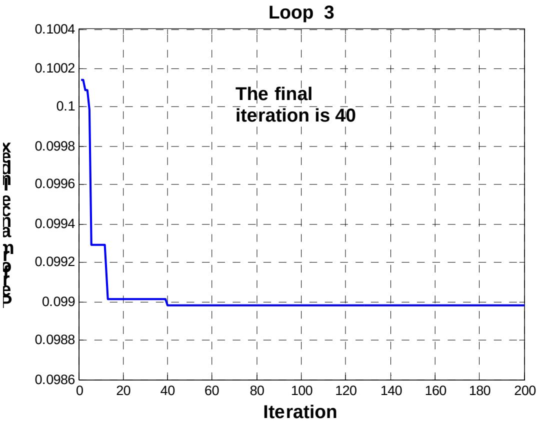

Fig.11.The performance index in the third loop

5. Conclusions

This paper briefly described development of a novel intelligent method for optimizing the parameters of a controller for a system which is interacted by an environment. The method developed was based on the extension of the conventional DARLA method. In order to decrease the number of iterations, the extended DARLA method has been applied to an *n*-dimensional space in which *n* stands for system parameters. By using matrix calculation, the speed of convergence can be increased resulting system improvement in real time. This method thus operates better in a system which is not robust and it can be used as a better alternative to conventional methods for determining the parameters of a PID controller . The method can also be used to optimise various case studies.

6. Acknowledgements

The authors wish to thank the Fuzzy Systems and Applications Center of Excellence , Shahid Bahonar University of Kerman , Kerman , Iran .

7.References

[1] G. Tesauro et al., 2006: A Hybrid Reinforcement Learning Approach to Autonomic Resource Allocation, *Proc. Int'l Conf. Autonomic Computing* (ICAC), IEEE CS Press, pp. 65–3.

[2] Imanipour, A., Ahmadabadi, M.N., Araabi, B.N., Asadpour, M., and Siegwart, R., 2006: Knowledge-based extraction of area of expertise for cooperation in learning, in *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst. (IROS)*, Beijing, China, pp.3700–3705.

[3]Tesauro, G., 2007: Reinforcement Learning in Autonomic Computing ",*Published by the IEEE Computer Society , IEEE INTERNET COMPUTING*.pp.22-30.

[4] M. N. Howell, G. P. Frost, T. J.Gordon and Q. H. Wu, "Continuous Action Reinforcement Learning Applied to Vehicle Suspension Control," *Mechatronics*, pp. 263-276, 1997.

[5] Ziegler, J. G., & Nichols, N. B. (1942). Optimum setting for Automatic Controllers. *Transactions of ASME*, *64*, 759-768.

[6] Astron, K. J., & Hagglund, T. (1995). *PID controllers: Theory, design, and tuning*. Research Triangle Park, NC, USA: ISA.