Control of Nonlinear Systems via Temporal Difference Learning Based Intelligent Controller and Context Reasoning

Javad Abdi[1] , Farzan Rashidi[2]

1. Center of Excellence for Control and Intelligent Processing, Dep. of Electrical and Computer Eng. University of Tehran, Tehran, Iran
2. Control Research Department, Engineering Research Institute, Tehran, Iran

Abstract:

Modeling emotions has attracted much attention in recent years, both in cognitive psychology and design of artificial systems. Far from being a negative factor in decision-making, emotions have shown to be a strong faculty for making fast satisfying decisions. In this paper, we have adapted a computational model based on the limbic system in the mammalian brain for control engineering applications.

Learning in this model based on Temporal Difference Learning. We applied the proposed controller (termed BELBIC) for a simple model of a submarine. The model was supposed to reach the desired depth underwater. Our results demonstrate excellent control action, disturbance handling and system parameter robustness for TDBELBIC.

The proposal method, regarding the present conditions, the system action in the part and the controlling aims, can control the system in a way that these objectives are attained in the least amount of time and the best way.

MODELLING

Motivated by the success in functional modeling of emotions in control engineering applications [15,29,30], the main purpose of this research is to use a structural model based on the limbic system of mammalian brain, for decision making and control engineering applications. We have adopted a network model developed by Moren and Balkenius [21], as a computational model that mimics amygdala, orbitofrontal cortex, thalamus, sensory input cortex and generally, those parts of the brain thought responsible for processing emotions. There are two approaches to intelligent and cognitive control. In the indirect approach, the intelligent system is utilized for tuning the parameters of the controller. We have adopted the second, so called direct approach, where the intelligent system, in our case the computational model termed TDBELBIC, is used as the controller block. The model is illustrated in figure 1. TDBELBIC is essentially an action generation mechanism based on sensory inputs and emotional cues. In general, these can be vector valued, although in the benchmarks discussed in this paper for the sake of illustration, one sensory input and one emotional signal (stress) have been considered. The emotional learning occurs mainly in amygdala. The learning rule of amygdala is given in formula (1).

$$\Delta G_a = k_1 . \max(0, EC - A) \qquad (1)$$

where G_a is the gain in amygdala connection, k_1 is the learning step in amygdala and EC and A are the values of emotional cue function and amygdala output at each time. The term $\max$ in the formula (1) is for making the learning changes monotonic, implying that the amygdala gain can never be decreased. This rule is for modeling the incapability of unlearning the emotion signal (and consequently, emotional action), previously learned in the amygdala [21,23]. Similarly, the learning rule in orbitofrontal cortex is shown in formula (2).

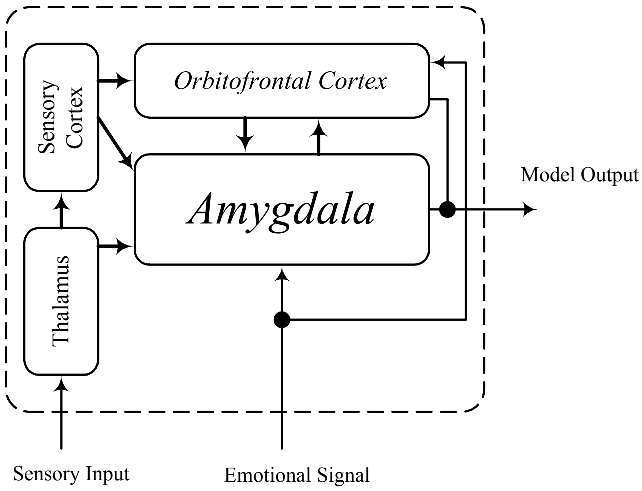

Figure 1- The abstract structure of TDBELBIC

$$\Delta G_o = k_2.(MO - EC) \quad (2)$$

where G_o is the gain in orbitofrontal connection, k_2 is the learning step in orbitofrontal cortex and MO is the output of the whole model, where it can be calculated as formula (3):

$$MO = A - O \quad (3)$$

in which, O represents the output of orbitofrontal cortex.

In fact, by receiving the sensory input S, the model calculates the internal signals of amygdala and orbitofrontal cortex by the relations in (4) and (5) and eventually yields the output.

$$A = G_a.S \quad (4)$$

$$O = G_o.S \quad (5)$$

Since amygdala does not have the capability to unlearn any emotional response that it ever learned, inhibition of any inappropriate response is the duty of orbitofrontal cortex.

IMPLEMENTAION

Controllers based on emotional learning have shown very good robustness and uncertainty handling properties [29,30], while being simple and easily implementable. To utilize our version of the Moren-Balkenius model as a controller, we note that it essentially converts two sets of inputs into the decision signal as its output. We have implemented a closed loop configuration using this block (termed TDBELBIC) in the feed forward loop of the total system in an appropriate manner so that the input signals have the proper interpretations. The block implicitly implemented the critic, the learning algorithm and the action selection mechanism used in functional implementations of emotionally based (or generally reinforcement learning based) controllers, all at the same time [15,29,30]. The structure of the control circuit we implemented in our study is illustrated in figure 2. The functions we used in emotional cue and sensory input blocks are given in (6) and (7),

$$EC = W_1.e + W_2.CO \qquad (6)$$

$$SI = W_3.PO + W_4.\dot{PO} \qquad (7)$$

where EC, CO, SI and PO are emotional cue, controller output, sensory input and plant output and the W_1 through W_4 are the gains must tuned for designing a satisfactory controller.

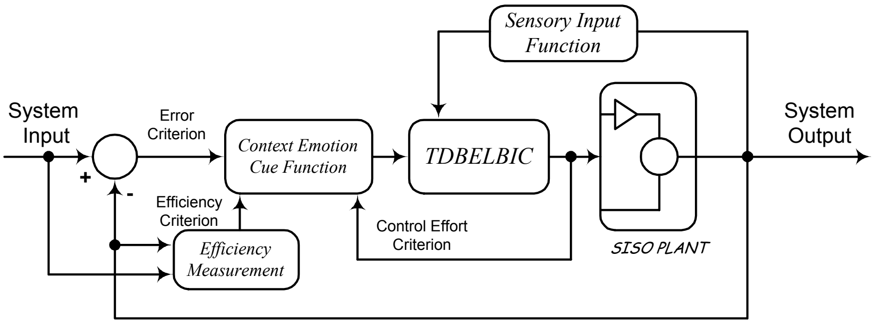

Figure 2 – Control system configuration using TDBELBIC

SIMULATIONS

We confirmed the capability of TDBELBIC by performing some simulations. It must be mentioned that in the all simulations outlined below, we implemented the set-point control strategy with the desired value of 1. The descriptions of simulations are given below:

LINEAR SISO SYSTEM: SUBMARINE MODEL

In this simulation, we considered a simple model of a submarine. The model was supposed to reach the desired depth underwater. The quantitative model is represented via (8).

$$G(s) = \frac{0.1(s+1)^2}{s(s^2+0.09)} = \frac{0.1s^2+0.2s+0.1}{s^3+0.09s} \qquad (8)$$

We implemented the control circuits in MATLAB SIMULINK package. The output of the system with a simple feedback and the output of the system with a TDBELBIC controller are given in figure 3.

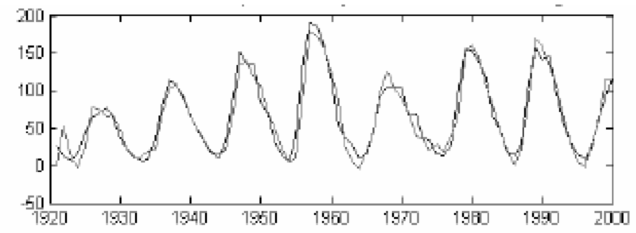

figure 3: The response of TDBELBIC controller

$TD(\lambda)$ Learning

Most of new learning algorithms like reinforcement learning, Q-learning and the method of temporal differences are characterized by their fast computation and in some cases lower error in comparison with the classical learning methods. Fast training is a notable consideration in some control applications. However, in prediction applications, two more desired characteristics of a good predictor are accuracy and low computational complexity.

In reinforcement learning, there is no teacher available to give the correct output for each training example, which is called unsupervised Learning. The output produced by the learning agent is fed to the environment and a scalar reinforcement value (reward) is returned. The learning agent tries to adjust itself to maximize the reward. [1][2]

Often that the actions taken by the learning agent to produce an output will affect not only the immediate reward but also the subsequent ones. In this case, the immediate reward only reflects partial information about the action. It is called delayed-reward. [2][3]

Temporal difference (TD) learning is a type of reinforcement learning for solving delayed-reward prediction problems. Unlike supervised learning, which measures error between each prediction and target, TD uses the difference of two successive predictions to learn that is Multi Step Prediction. The advantage of TD learning is that it can update weights incrementally and converge to a solution faster. [4]

In a delay-reward prediction problem, the observation-outcome sequence has the form $x_1, x_2, x_3, ..., x_m, z$ where each x_t is an observation vector available at time $t, 1 \le t \le m$ and z is the outcome of the sequence. For each observation, the learning agent makes a prediction of z, forming a sequence: $P_1, P_2, P_3, ..., P_m$.

Assuming the learning agent is an artificial neural network, update for a weight w of the network with the classical gradient descent update rule for supervised learning is:

$$\Delta w = -\alpha \nabla_w E = -\alpha \sum_{t=1}^{m} (P_t - z) \nabla_w P_t \qquad (9)$$

Where α is the learning rate and $\nabla_w E$ is the gradient vector, $\frac{\partial E}{\partial w}$ of the mean square error function:

$$E = \frac{1}{2} \sum_{t=1}^{m} (P_t - z)^2 \qquad (10)$$

In [3], Sutton derived the incremental updating rule for equation (9):

$$\Delta w_t = \alpha (P_{t+1} - P_t) \sum_{k=1}^{t} \nabla_w P_k \qquad (11)$$

For $t = 1,2,\ldots,m$ where $P_{m+1} \stackrel{def}{=} z$

To emphasize more recent predictions, an exponential factor λ is multiplied to the gradient term:

$$\Delta w_t = \alpha(P_{t+1} - P_t)\sum_{k=1}^{t} \lambda^{t-k} \nabla_w P_k \qquad (12)$$

Where $0 \le \lambda \le 1$

This results in a family of learning rules, $TD(\lambda)$, with constant values of λ.
But there are 2 special cases:
First, when $\lambda = 1$, Eq. (12) falls back to Eq. (11), which produces the same training result as the supervised learning in Eq. (9). Second, when $\lambda = 0$, since $0^0 = 1$, Eq. (12) becomes

$$\Delta w_t = \alpha(P_{t+1} - P_t)\nabla_w P_k \qquad (13)$$

I can extended the Eq. (13) for BELBIC and made Eq. (14) for TDBELBIC.

$$\Delta G_{Ot} = \alpha(z - P_t)\nabla G_O P_t \qquad (14)$$

Which has a similar form as Eq. (9). So the same training algorithm for supervised learning can be used for $TD(0)$.

Conclusion:

In figure (3), you can observe the results of simulating the diagram block figure (2). The results, based on temporal difference learning, are compared to Orbitofrontal Cortex learning in a shared TDBELBIC structure. The outcomes suggest that temporal difference based learning in faster than Orbitofrontal Cortex learning. But faster learning is increased for maximum overshoot. Both of the learning is incremental, however, their memory output signals are presented in figure (4), (5). The increase rates represent their learning speed.
Paying attention to the achievements in the emotional controls founded a computational model, based on the Limbic system, for mammals' brain via time series learning. The paper tried to develop this method for answering more complicated issues and achieving difficult goals.

To do this, the ability of the learning module the emotional controller, was increased achieving based a brain computational model means of temporal difference learning for credit assignment. Temporal difference learning, has easier computations because of using it's own experience. The methods resemble human behavioral learning.

BEBIC Output

TDBEBIC Output

BELBIC Emotional Signal

TDBELBIC Emotional Signal

Amygdala BELBIC Output

Amygdala TDBELBIC Output

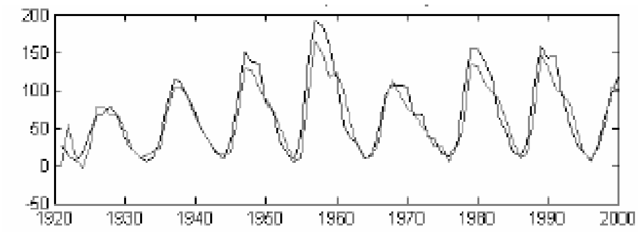

Figure (4): Comparison of BELBIC and OFC Learning with BELBIC and TD Learning

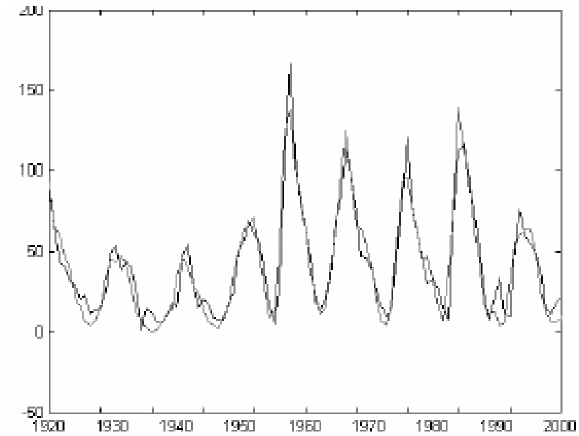

Figure (5): Comparison of BELBIC and with TDBELBIC Memory

References:

[1] J. Abdi, C. Lucas, "Survey about the Application of Temporal Difference and Reinforcement Learning in control", MSc *Dissertation, Department of Computer and Electrical Engineering,* University of Tehran , IRAN, 2002

[2] David Eby, R. C. Averill, William F. Punch III, Erik D. Goodman "Evaluation of Injection Island GA Performance on Flywheel Design Optimization", *January 15, 1998, Proceedings, Third Conference on Adaptive Computing in Design and Manufacturing, Plymouth, England, April, 1998, Springer Verlag*

[3] R. S. Sutton, "Learning to Predict by the Methods of Temporal Differences", *Machine Learning, 3:9-44,1988*

[4] R. S. Sutton, "Generalization in Reinforcement Learning: Successful Examples Using Sparse Coarse Coding*", To Appear in Advances in Neural Information Processing Systems 8, 1995*

[5] J. H. Steele "Approximation and Validation of Models with Uncertainty: a closed-loop perspective" *Ph.D. thesis, August 2001, Control Group Department of Engineering, University of Cambridge*

[6] C. Lucas, D. Shahmirzadi and N. Sheikholeslami, "INTRODUCING A CONTROLLER BASED ON BRAIN EMOTIONAL LEARNING ALGORITHM: BELBIC", *Submitted to International of Intelligent Automation and Soft computing (Autosoft), August 2002*

[7] C. Lucas, B.N. Arabi, D. Shahmirzadi and O. Namaki, "Speed Control of a Switched Reluctance Motor using BELBIC", *Submitted to IFAC Journal of Control Engineering Practice, November 2002*

[8] A. Crameri, S. A. Raillard, E. Bermudez, W. P. C. Stemmer "DNA Shuffling of a Family of Genes from Diverse Species Accelerates Directed Evolution" *Nature, 391(6664), PP: 288-291, January 1998*

[9] H. V. D. Parunak "Go to The Ant: Engineering Principles From Natural Multi-Agent Systems" *Annals Of Operations Research, 75 (Special Issue On Artificial Intelligence And Management Science), PP: 69-101, 1997*

[10] M. Sipper, M. Tomassini "Convergence to Uniformity In A Cellular Automaton Via Local Coevolution" *International Journal Of Modern Physics C, Vol. 8, No. 5, PP: 1013-1024, 1997*

[11] M. Toda "The Urge Theory of Emotion and Cognition" *Emotion and urges, SCCS Technical Report, No.93-1-01, 1993*
[12] R. Neese "Emotional Disorders in Evolutionary Perspective" *British Journal Of Medical Psychology (1998), 71, PP: 397-415*
[13] J. D. Greene, R. B. Sommerville, L. E. Nystorm, J. M. Darley, J. D. Cohen "An FMRI Investigation of Emotional Engagement in Moral Judgment" *Science, Vol. 293, 14 Sep. 2001, PP: 2105-2108*
[14] R. D. Smallwood, E. J. Sondik " The Optimal Control of Partially Observable Markov Processes Over a Finite Horizon" *Operational Research, (21), 1973, PP: 1071-1088*
[15] J. R. Jang "ANFIS: Adaptive-Network-Based Fuzzy Inference System" *IEEE Transaction on Systems, Man and Cybernetics, Vol. 23, No. 3, 1993*
[16] S. Hofmeyr, S. Forrest "Architecture for an Artificial Immune System" *Evolutionary Computation Journal, 2000*
[17] X. Xu, H-G He, D. Hu "Efficient Reinforcement Learning Using Recursive Least-Squares Methods" *Journal of Artificial Intelligence Research 16, PP: 259-292, 2002*
[18] C. Lucas, D. Shahmirzadi, M. Biglarbegian "Co-evolutionary Approach for Graph-Coloring Problem" *Accepted to Amirkabir University of Technology Press, April 2002*
[19] C. Lucas, D. Shahmirzadi "A Novel Interpolative Fuzzy Inference Procedure Using Least Square Principle" *Accepted to Journal of Control and intelligent systems, December 2002*
[20] D. Shahmirzadi, C. Lucas, M. Nikkhah Bahrami, H. Ghafoorifard "A FEM-Based Quasi-Static Neuro-Model for Acoustic Noise In Switch Reluctance Motors" *Accepted to Journal Of Computational Acoustics, June 2002*
[21] M. H. Voigt "Fuzzy evolutionary algorithms" *June 1992, Technical Report 92-038, International Computer Science Institute (ICSI), 1947 Center Street, Suite 600, Berkeley, CA, 94704*
[22] A. Bergman, W. Burgard, A. Hemker "Adjusting parameters of genetic algorithms by fuzzy control rules" *New Computer Techniques in Physics Research III, pages 235--240. World Scientific Press, 1994*
[23] J. Chen, E. Antipov, B. Lemieux, W. Cedeno, D. H. Wood "DNA Computing Implementing Genetic Algorithms" *Preliminary Proceedings DIMACS Workshop On Evolution As Computation, PP: 39-49 , DIMACS, Piscataway NJ, January 1999*
[24] J. S. Bay "Behavior Learning in Large Homogeneous Populations of Robots" *IASTED International Conference on Artificial Intelligence and Soft Computing, PP: 137-140, July 1997*
[25] K. Inoue, K. Kawabata, and H. Kobayashi "On a Decision Making System with Emotion" *5th IEEE International Workshop on Robot and Human Communication, 1996, PP: 461-465*
[26] H. Jeschke "Fuzzy Multi Objective Decision Making on Modeled VLSI Architecture Concepts" *International Symposium on Circuits and Systems 1998, June 1-3, Monterey, CA, USA*
[27] J. Moren, C. Balkenius "A Computational Model of Emotional Learning in The Amygdale" *From animals to animals 6: Proceedings of the 6th International conference on the simulation of adaptive behavior, Cambridge, Mass., 2000. The MIT Press*
[28] K. Miyazaki, N. Araki, E. Mogi, T. Kobayashi, Y. Shigematsu, M. Ichikawa, G. Matsumoto "Brain Learning Control Representation In Nucleus Accumbency" *1998 Second International Conference On Knowledge-Based Intelligent Electronic Systems, PP: 21-23, April 1998, Australia*
[29] C. Balkenius, J. Moren "A Computational Model Of Emotional Conditioning In The Brain" *Proceeding of the workshop on Grounding Emotions in Adaptive Systems, Zurich, 1998*
[30] A. A. Savinov "An Algorithm for Induction of Possibilities Set-Valued Rules by Finding Prime Disjunctions" *4th On-line World Conference on Soft Computing in Industrial Applications (WSC4), September 21-30, 1999*
[31] D. Shahmirzadi "Proposing A Genetic Algorithm For Solving The Map-Coloring Problem" *Proceeding of 4th student conference on intelligent systems, March 5-7, 2002, K.N. Toosi University Of Technology, Tehran, Iran*
[32] D. Shahmirzadi, C. Lucas, M. Nikkhah Bahrami, H. Ghafoorifard "A Computational Approach To Acoustic Noise In Switch Reluctance Drives" *Proceeding of 3rd International Conference on Mathematical and Computational Applications, ICMCA'2002, September 4-6, 2002, Konya, Turkey*
[33] M. Fatourechi, C. Lucas, A. Khaki Sedigh "An Agent-based Approach to Multivariable Control" *Proceedings of IASTED International Conference on Artificial Intelligence and Applications", PP: 376-381,Sept 4-7, 2001, Marbella, Spain*
[34] C. Lucas, S. A. Jazbi, M. Fatourechi, M. Farshad "Cognitive Action Selection with Neuro-controllers" *Third Irano-Armenian Workshop on Neural Networks, August 2000, Yerevan, Armenia*

[35] M. Fatourechi, C. Lucas, A. Khaki Sedigh "Reducing Control Effort by means of Emotional Learning" *Proceedings of 9th Iranian Conference on Electrical Engineering, (ICEE2001), PP: 41-1 to 41-8, May, 2001, Tehran, Iran*
[36] M. Fatourechi, C. Lucas, A. Khaki Sedigh "Reduction of Maximum Overshoot by means of Emotional Learning" *Proceedings of 6th Annual CSI Computer Conference, PP: 460-467, February, 2001, Isfahan, Iran*
[37] A. Ott "The Rule Set Based Access Control (RSBAC)" *The 8th International Linux Congress, Enschede, November 28-30, 2001*
[38] D. Shahmirzadi, C. Lucas, M. Farshad, R. Pedrami "Training Neural Network For Modeling The Acoustic Behavior Of Switch Reluctance Motor" *Neuroscience events International conference, proceeding of 4th Irano-Armenian workshop on neural networks, May 20-22, 2002, School Of Intelligent Systems, Institute for studies in theoretical physics and mathematics, Tehran, Iran*
[39] E. J. Davis, G. Kendall "An Investigation, using Co-Evolution, to evolve an Aware Player" *Proceedings of Congress on Evolutionary Computation (CEC2002), PP: 1408-1413, Hilton Hawaiian Village Hotel, Honolulu, Hawaii, May 12-17, 2002*
[40] P.Cichosz, "Truncating Temporal Differences: On the Efficient Implementation of $TD(\lambda)$ for Reinforcement Learning", *Journal of Artificial Intelligence Research, 2:287-318, 1995*