American Journal of Intelligent Systems 2012, 2(7): 184-190
DOI: 10.5923/j.ajis.20120207.05

RLIDS Based Goal Improvement in Emotional Control

Amin Amiri Tehrani Zade*, Iman Esmaili Paeen Afrakoti, Saeed Bagheri Shuraki

Artificial Creature Lab, Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran

Abstract In this paper, a goal improvement approach based on experiences for handling external disturbances is proposed. The suggested method uses reinforcement learning by an actor-critic mechanism in order to adapt the system to disturbance and uncertainty. Goal improvement is achieved via IDS. IDS is a powerful method in information-handling processes. It has a fast convergence and simple underlying, away from complex formulas. Temporal Difference (TD) learning is utilized in updating both Stress-Evaluator and the goal of emotional controller. As a benchmark to evaluate our method, inverted pendulum and simple submarine systems are used. Results show that the proposed method responds better to external tackling disturbances in comparison with the former fixed goal model.

Keywords Goal, BELBIC, IDS, Fuzzy Surface, Learning

1. Introduction

What is a Goal? This is the question that must be answered before going any further. The term "Goal" is defined as the cognitive representation of a desirable state which affects evaluation, emotion and behaviour. This definition has been put forward by psychology researchers[1-3]. Furthermore, a goal requires a variety of activities which enable us to reach a desirable state[4]. Goals have positivity nature which has motivation force. They also have primary reason that influence our behaviour due to their positivity natures[5]. Although Goal is defined as a desirable state which has positivity nature, the reason for which goals become positive is not clear. A goal might become positive and desirable either consciously and intentionally or subconsciously and involuntarily, e.g. repeated pairing of a given activity and the consequent reward[6].

To develop algorithms in engineering and decision making systems based on psychological and biological mechanisms is a promising area of research[7-9]. The challenging part of any psychological or biological system development is its learning necessity to adapt itself to random incidents inherent in the environment[7, 9].

Goals, as defined in engineering problems, are the performance functions that continuously evaluate the responses of environment. For instance, goals can have a critic function in critic based fuzzy controller[10], a fitness function in evolutionary algorithm[11] or they can function as an emotional cue in Brain Emotional Learning based Intelligent Controller (BELBIC)[12].

* Corresponding author:
Amin.amiriteh@gmail.com (Amin Amiri Tehrani Zade)
Published online at http://journal.sapub.org/ajis

Goal has a duty of directing the learning system to the desirable state. If the environment is corrupted by a variety of disturbances that cannot be predicted from the outset, adapting the parameter of learning agent is inevitable.

Up to now there hasn't been sufficient studies conducted on adapting goals in order to improve the performance of system in tackling disturbance for emotional control. Garmsiri et. Al[13] have proposed a fuzzy tuning system for parameter improvement. They used human knowledge and experience to extract fuzzy rules. Rouhani et. Al[14] have used expert knowledge to form a relationship between error and its emotional cue. In addition, a learning mechanism for attention control using sparse-super critic has been proposed in[9]. In this method, super critic sends a punishment signal sparsely in order to learn the degree of importance of each local critic assessment.

As noted above, one of the critic based learning systems is BELBIC. BELBIC was first introduced in[12]. Its structure is based on Moren's research in the field of emotional learning. Moren et Al[15] have proposed a computational model of limbic system of mammalian brain which is responsible for emotional processes. BELBIC is a model free controller with fast learning ability. Excellent performance of BELBIC in confronting disturbances has made it favourable in several control and decision making applications[12, 16-17].

In model free controllers, learning may cause instability. This drawback, especially for a plant with an unstable nature, has been a challenging problem for researchers in the area of emotional control. Roshtkhari et. Al[16] cope with this problem by an imitative learning technique. Their method is based on designing imitative phases of learning to stabilize the inverted pendulum.

In this paper, we introduce a new mechanism for learning and improving goal in BELBIC. In order to improve a goal,

we make use of a mechanism like actor-critic. Actor-critic is a temporal difference based learning method that uses an interaction mechanism to learn proper action in a state-action space. Comprehensive details of this method can be found in[18]. The main difference of our method with other methods is that critic in our approach is an actor, and Stress-Evaluator is a critic in actor-critic structure.

State-stress value planes are modeled as RLIDS in[19]. Ink drop spread (IDS) is employed as RLIDS engine. IDS is a fuzzy modeling algorithm, which expresses multi-input - single-output system as a fuzzy combination of several single-input-single-output systems[20-22]. In our algorithm, initial goal and its update rule is designed using IDS method.

The paper is organized in 6 sections. In section 2 BELBIC is introduced. Section 3 is dedicated to IDS and RLIDS methods. Section 4 is devoted to our proposed method. In section 5, simulation results are demonstrated. Finally, section 6 concludes the paper.

2. BELBIC

BELBIC is a learning controller based on the computational model of Moren's limbic structure[15]. This structure is a simple model of the main parts in limbic system i.e., amygdale, orbitofrontal cortex, thalamus and sensory cortex which is shown in figure 1.

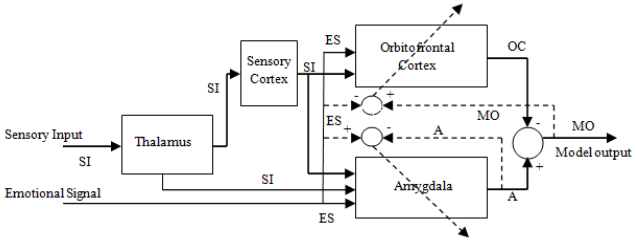

Figure 1. simple structure of limbic system[23]

As we can see in figure 1, sensory input is first processed in thalamus. After this stage of pre-processing of input signals, they are sent to sensory cortex and amygdale units. Sensory cortex has the duty of discriminating the coarse output from thalamus. Then, this filtered signal is sent to amygdale and orbitofrontal cortex. Amygdale is a small structure in the medial temporal lobe of brain which is responsible for the emotional evaluation of stimuli[16]. Another main component in limbic system is orbitofrontal cortex that has the duty of inhibiting inappropriate responses from amygdale.

If several sensory inputs are required for controller design, thalamus output will be the max of these sensory inputs ,as shown in formula (1), otherwise single sensory input will directly be sent to amygdale.

$$S_{th} = \max(S_i) \tag{1}$$

As we can see in figure 1, output of thalamus is not entered in orbitofrontal cortex. This means that this signal is not obstructed.

Output of amygdala and orbitofrontal cortex is computed in formula (2), (3) and (4).

$$A_i = S_i V_i \tag{2}$$

$$A_{th} = S_{th} V_{th} \tag{3}$$

$$O_i = S_i W_i \tag{4}$$

Based on output of Amygdala and Orbitofrontal cortex, model output is computed using formula (5).

$$E = \sum A_i - \sum O_i + A_{th} \tag{5}$$

Learning in Amygdala and Orbitofrontal cortex is also defined through formula (6), (7) and (8) .

$$\Delta V_{th} = \alpha.\max(0, S_{th}(stress - (\sum A_i + S_{th} V_{th}))) \tag{6}$$

$$\Delta V_i = \alpha.\max(0, S_i(stress - (\sum A_i + S_{th} V_{th}))) \tag{7}$$

$$\Delta W_i = \beta.S_i(E - stress) \tag{8}$$

In above formulas, α and β are learning rates that determine the rate at which they affect emotional cues into model output. As we can see in formula (6) and (7), amygdala have the function of directing output in order for the association to be made between sensory input and emotional cue signal. Also ,as seen in formula (8), learning in orbitofrontal cortex will cause output inhibition if expected signal mismatches that of reinforcement.

As noted in section 1, BELBIC is a critic based controller whose critic is in emotional cue type e.g., anger, fear, happiness, sadness or stress. We accept stress as an emotional cue signal and design our mechanism based on this concept.

3. Ink Drop Spread

Ink Drop Spread (IDS) is a fuzzy operator that has the function of information propagation in order to extract total information about the behaviour of data in the universe of discourse. In other words, IDS is a fuzzy interpolation operator. It is capable of deriving smooth curve from input data. In this method, Interpolation can be done by the use of Pyramid as a three dimensional fuzzy membership function of a data point and its neighbouring points , as shown in figure 2.

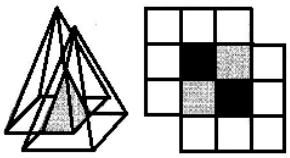

Figure 2. pyramid membership functioning as the IDS operator and its correlation with neighbouring points

IDS is the heart of Active Learning Method (ALM), i.e. the modeling technique in the field of fuzzy modeling. In this modeling technique, a multiple inputs-single output (MISO) system is broken into several single input-single output (SISO) systems. Useful information is extracted from these simple single-input systems, and output is inferred from the

combination rule on this information. Figure 3 shows the flowchart of ALM algorithm.

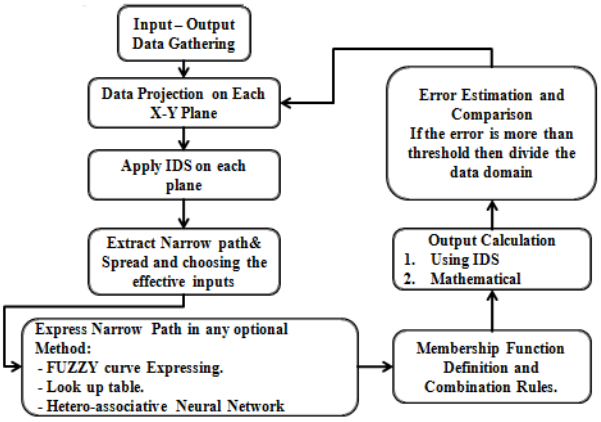

Figure 3. flowchart of ALM Algorithm

The approach in this method is to divide each of the inputs into pre-determined fuzzy partitions first in order to get better accuracy of data in SISO systems. Then, each piece of data is mapped into one single-input-single-output plane in accordance with its partition. IDS operator is employed afterward, and system behaviour and inference rules are extracted.

Two concepts are extracted from an IDS plane: "Narrow path" and "Spread of data". Each narrow path shows the behaviour of an output relative to the input. Spread of data pointing around this path shows the importance degree of that input in overall system behaviour[24].

This modelling technique is inspired by the concept of active learning in the human brain. But, What is meant by active learning? Active learning is learning through a recursive approach in which the brain actively receives a question; then it searches to find the best answer and forgets the previous answer that was incorrect. In other words, this learning system itself generates a new pattern of learning that has never been experienced before. This approach is called Active Learning in psychology[25].

While other well-known modelling methods suffer from computational complexity, this method is far from this problem. This simplicity is a result of imitating human learning. Moreover, when we are seeking implementation of a human inspired method, it is more rational to use methods and algorithms that have similar bases as those of humans.

In RLIDS, a reinforcement learning method based on actor-critic system similar to Generalized Approximate reasoning based Intelligent Control (GARIC) has been utilized. In this method, IDS is used as the main engine in updating the actor and critic.

4. Proposed Algorithm

As mentioned in section 1, one of the goal improvement approaches in humans is the consequence of reward experience. Reward experience refers to getting a reward for an action we do. Therefore, a mechanism that evaluates our action in each state must exist. In addition to the above statement, because our aim is to enter the mechanism of goal improvement in the area of intelligent control, we have considered two main parts in our system: Stress-Evaluator and state-stress values. Stress-Evaluator evaluates stressful actions. The state-stress evaluates the memory structure like IDS planes that memorize past experiences and their goodness according to Stress-Evaluator signals. Block diagram and flowchart of proposed controller is illustrated in figure 4 and 5.

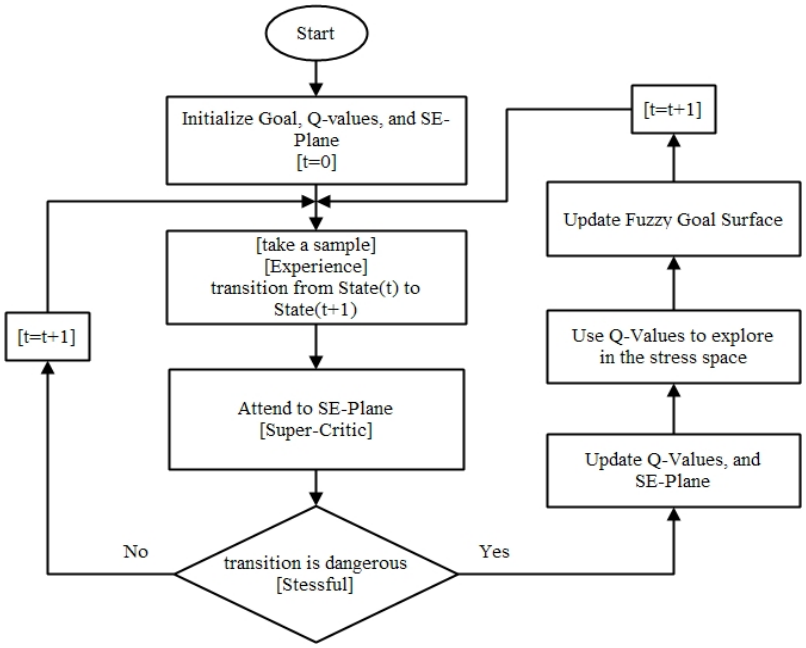

Figure 4. flowchart of Proposed Algorithm

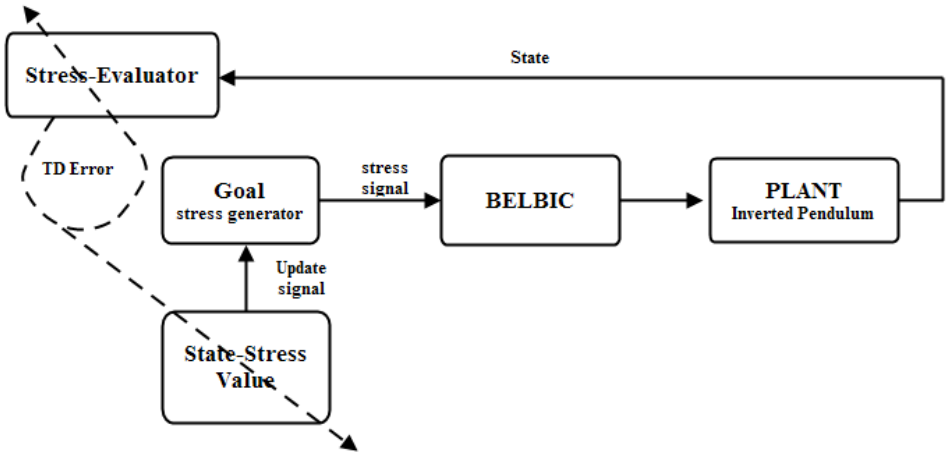

Figure 5. block diagram of proposed learning mechanism

4.1. Stress-Evaluator

Stress-Evaluator is a mechanism over the critic that helps it to improve for best interaction with unknown disturbances of the environment.

Stress-Evaluator in stress based critics is a metric for danger. The design of Stress-Evaluator is arbitrary for the designer and his objectives. One of the many choices that we use in this paper is Stress-Evaluator, as shown in figure 5.

In the design of Stress-Evaluator, two points must be considered:

- The larger the error or its derivative is, the larger the danger will be. Therefore in such situations, we must use our past experiences much more. Thus, the range of exploration is limited.
- We must bear in mind that in all actions, certain amount of stress is existent; therefore our aim is to bring danger to a limited range. If error and derivative of error are lower than the threshold, this action is good and no update is needed.

4.2. State-stress value planes

For exploration, we must save past experiences. These experiences are saved in the form of IDS planes which are inspired by active learning in human brain. In addition to experience, some extra patterns are created in accordance with the relationship among the gathered experiences.

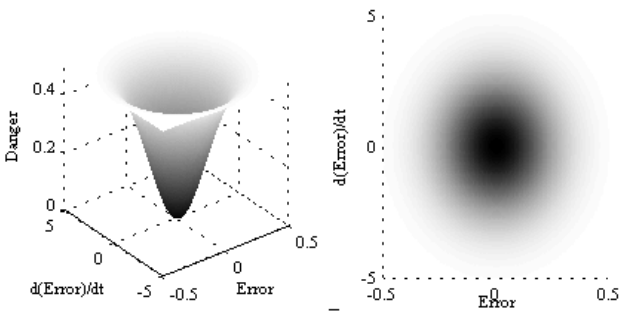

Figure 6. Stress-Evaluator Plane

We consider one IDS plane for each variable. Especially in our design, we used two IDS planes corresponding to error and its derivative. These state-stress value planes are used not only for storing values, but also for determining the amount of shift in fuzzy goal surfaces.

The amount of shift is calculated by

$$shift = stress_{selected} - stress_{actual} \quad (9)$$

In formula (9) $stress_{selected}$ is obtained by roulette wheel mechanism. $Stress_{actual}$ is the amount of stress determined by the goal fuzzy surface. First, values of all IDS planes are set to zero. After each experience, based on the evaluation of the Stress-Evaluator, these values are updated.

4.3. Learning Mechanism

Total structure of learning is in the type of actor-critic and is based on Temporal Difference learning. The main difference is that in our approach Stress-Evaluator criticizes the action generated based on stress signal in that state. The evaluation of Stress-Evaluator is in the form of stressful action. Stress evaluation is achieved through Temporal Difference, as shown in formula (10). In this formula δ is the amount of danger in transition from state in time (t-1) to state in time (t). SE is Stress-Evaluator plane.

$$\delta = \lambda[SE_{t-1}(error, \frac{d(error)}{dt}) - SE_t(error, \frac{d(error)}{dt})] \quad (10)$$

Base on δ, ink drop (ID) is defined as

$$ID = \delta.H \quad (11)$$

In this formula H is an IDS window which shows to what extent the neighbours must be affected. It can be a pyramid, a Gaussian window or something else. Also, ID in formula (11) represents ink drop that must be used for updating state-stress value planes according to formula (12).

$$\begin{cases} Q\_IDS_{state\_\mathrm{var}}(stress)|_t = \\ Q\_IDS_{state\_\mathrm{var}}(stress)|_{t-1} + ID \quad if \ |\delta| < threshold \\ Q\_IDS_{state\_\mathrm{var}}(stress)|_t = \\ Q\_IDS_{state\_\mathrm{var}}(stress)|_{t-1} - ID \quad if \ |\delta| > threshold \end{cases} \quad (12)$$

Furthermore, SE-plane is updated by formula (13).

$$SE_t(error, \frac{d(error)}{dt}) = SE_{t-1}(error, \frac{d(error)}{dt}) + \beta.ID \ \ if \ \ \delta < 0 \quad (13)$$

As seen in formula (13), in each transition from one state to the next one, the danger increases, and the danger of previous state increases too. This concept is inspired by human danger management. When humans, in one state, take an action according to goal signal, if that action is stressful and danger is increased, they try to change the goal in order to take an action with lower stressful risk. Furthermore, danger of that state increases in his Stress-Evaluator mechanism. This increase in initial state danger causes larger distance from danger, leading to better exploration.

4.4. Goal Updating Mechanism

The initial goal is modeled using IDS method. In this method, one IDS plane is used for each rule. Each plane is based on one variable. In our problem, we considered two variables i.e., error and derivative of error. Shifts that are obtained by state-stress value planes in formula (9) are imposed by a Gaussian function on narrow path of corresponding IDS plane of the same variable and its neighbours.

5. Results

5.1. Example1: Inverted Pendulum

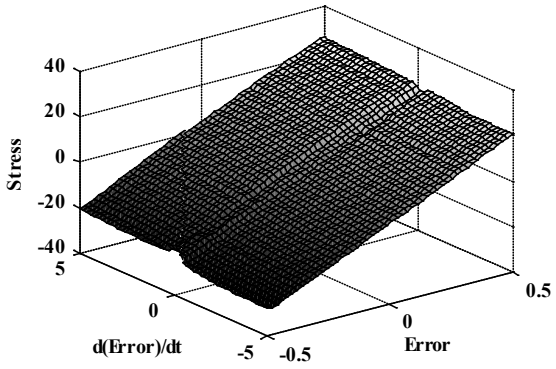

Figure 7. Initial stress fuzzy surface (Goal)

To validate the proposed controller, its results are compared with that of PID and BELBIC with fixed stress surface. In order to compare results rationally, first PID is designed using classical methods and then this PID is considered as sensory input in BELBIC. Stress generator fuzzy surface is designed as a critic to improve the performance of primary sensory input signal. Initial stress surface guaranties an initial stable response, not a proper response in confronting disturbance, as shown in figure 7. Then the proposed algorithm is employed on the basis of BELBIC controller. This means that improving the goal and BELBIC learning occur simultaneously.

In order to evaluate the performance of proposed algorithm in confronting disturbances, the sequence of random force disturbance by a Gaussian distribution with a frequency of 10, mean of 0 and variance of 6, is applied to the pendulum shown in figure 8.

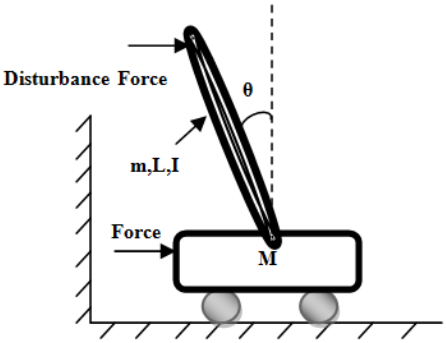

Figure 8. illustration of Inverted Pendulum and applied disturbance force

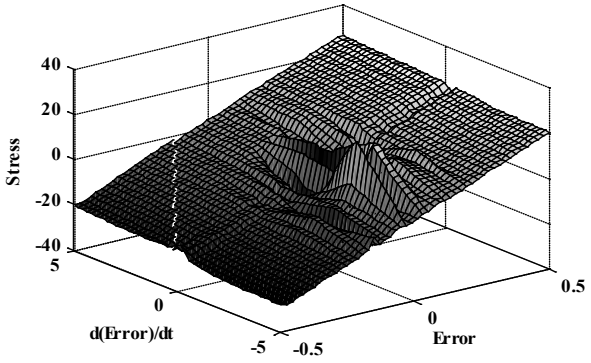

Figure 9. Stress surface after 40 iterations of learning under applied disturbance force

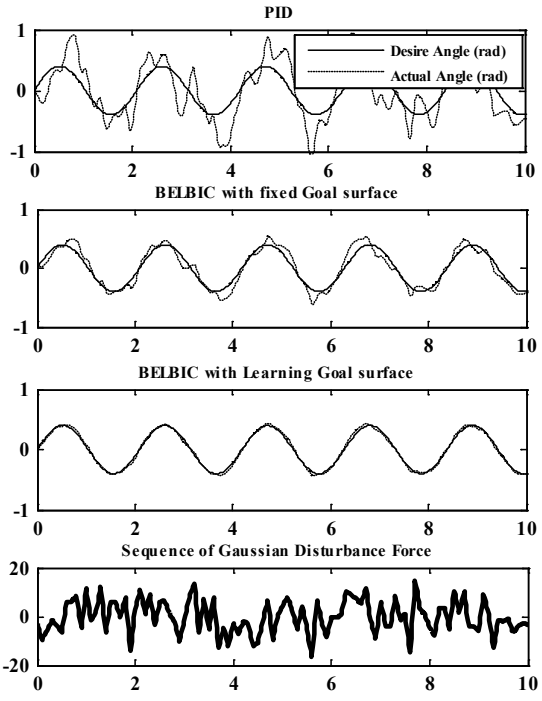

Figure 10. result of applying disturbance on inverted pendulum after 40 iterations of learning

Parameter of the inverted pendulum is considered as: M=0.5 Kg, m=0.2 Kg, L=0.6 m and I=0.006 Kgm^2.

Random disturbance was applied for 50 seconds and learning was achieved in 40 iterations in order to show convergence of algorithm due to the convergence of the level of danger in states of Stress-Evaluator Plane.

As it can be seen in figure 8, in the range of experienced error and derivative of error, stress fuzzy surface is updated in order to satisfy Stress-Evaluator signals and direct actions towards less stressful risks.

Results of 40 iterations of learning under the disturbance force together with sine desire angle are illustrated in figure 9.

As it can be seen in figure 9, proposed method truly detects stressful actions and finds the best state in order to eliminate stressful actions.

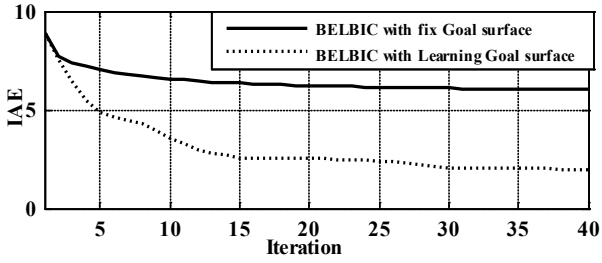

Figure 11. Evolution of IAE in 40 iterations of learning

In figure 10, IAE is the Integral Absolute Error for pendulum angle.

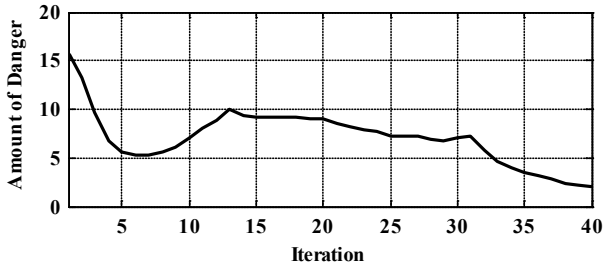

Figure 12. Amount of danger generated from Stress-Evaluator in process of control of inverted pendulum with learning goal in 40 iterations

As it can be seen in figure 11, each iteration of learning exploration and exploitation occurs simultaneously, but whenever there is more experience, the amount of exploitation increases. After a sufficient number of iterations, the amount of danger decreases gradually and the more it gets close to zero, the more convergent it becomes.

Table 1. Performance measure under disturbance force of figure 9

Controller	IAE(angle)	IACF
PID	11.7251	124.4301
BELBIC (Fix Goal)	5.9565	119.6301
BELBIC (Learning Goal)	1.8829	119.2643

In table 1, IACF is the Integral of Absolute values of Control Force.

5.2. Example2: Simple Submarine System

As another benchmark, a simple submarine system[12] is used. The transform function of this model is considered as:

$$\mathbf{G(s)} = \frac{\mathbf{0.1s^2+0.2s+0.1}}{\mathbf{s^3+0.09s}} \quad (14)$$

In order to evaluate our method in confronting varying dynamic, model of system is changed as:

$$\mathbf{G(s)} = \frac{\mathbf{0.1s^2+0.1s+0.2}}{\mathbf{1.1s^3+0.12s+1}} \quad (15)$$

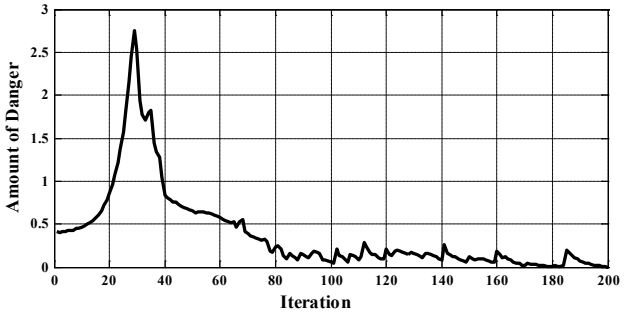

Figure 13. Amount of danger generated from Stress-Evaluator in control process of submarine system with learning goal in 200 iterations

Figure 12 illustrates danger management mechanism confronting the uncertainty due to varying dynamic in the submarine system. After approximately 80 iterations our model is experienced with uncertainty in system and danger is faded.

Table 2. Performance measure of submarine system response to step reference signal under varying dynamic as in formula 15

Controller	POS	Rise Time	Settling Time	S-S Error(%)
PID	22	0.33	11.31	-1.5
BELBIC (Learning Goal)	4	0.09	0.12	-0.2

As it is seen in table 1 and table 2, proposed algorithm has a good performance in improving the goal, which is stress in BELBIC controller.

6. Conclusions

In this paper, the main aim was to enter the mechanism of human goal improvement in the field of intelligent control. In BELBIC, controller goal is directed using the stress generated in emotional cue signal. Therefore, we used this concept as the goal and implemented the proposed mechanism in order to improve it in getting experience from the environment so after some experiences its performance improves.

In order to initialize and update goal surface, IDS method is used and the structure of actor-critic is utilized for learning. Stress-Evaluator is used as the critic of stress generator fuzzy surface which represents our goal of control.

Results show that our approach truly detects stressful actions and directs the goal towards increasing the performance of the overall controller.

REFERENCES

[1] M. S. Charles S. Carver, Attention and self-regulation: a control-theory approach to human behavior. New York: Springer, 1981.

[2] P. M. Gollwitzer, Moskowitz, G. B., Goal effect on thought and behavior. New York: Guilford Press, 1996.

[3] E. T. Higgins, Kruglanski, A. W., Motivational science: Social and personality perspectives. Philadelphia: Psychology Press, 2000.

[4] D. A. Norman, "Categorization of action slips," Psychological Review, vol. 88, pp. 1-15, 1981.

[5] J. W. Atkinson, Strength and motivation and efficiency of performance. New York: Wiley, 1974.

[6] R. Custers, Aarts, H., "Positive Affect as Implicit Motivator: On the Nonconscious Operation of Behavioral Goals," Personality and Social Psychology, vol. 89(2), pp. 129-142, 2005.

[7] G. Klein, A recognition-primed decision model of rapid decision making, In Decision Making in actions: models and methods. New Jersey, United States: Albex Publishing Corp, 1993.

[8] A. Nowroozi, et al., "A general computational recognition primed decision model with multi-agent rescue simulation benchmark," Information Sciences, 2011.

[9] A. Arami, et al., "Attention to multiple local critics in decision making and control," Expert Systems with Applications, 2011.

[10] A. Fakhrazari and M. Boroushaki, "Adaptive critic-based neurofuzzy controller for the steam generator water level," IEEE Transactions on Nuclear Science, vol. 55, pp. 1678-1685, 2008.

[11] J. E. S. A. E. Erben, Introduction to Evolutionary Computing: Springer, 2003.

[12] C. Lucas, et al., "Introducing BELBIC: Brain emotional learning based intelligent controller," Intelligent Automation and Soft Computing, vol. 10, pp. 11-22, 2004.

[13] N. Garmsiri and F. Najafi, "Fuzzy tuning of brain emotional learning based intelligent controllers," Jinan, 2010, pp. 5296-5301.

[14] H. Rouhani, et al., "Brain emotional learning based intelligent controller applied to neurofuzzy model of micro-heat exchanger," Expert Systems with Applications, vol. 32, pp. 911-918, 2007.

[15] J. Moren and C. Balkenius, "A Computational Model of Emotional Learning in the Amygdala: From animals to animals," presented at the Proc. of 6th International Conference on the Simulation of Adaptive Behavior, cambridge, 2000.

[16] M. Javan-Roshtkhari, et al., "Emotional control of inverted pendulum system: A soft switching from imitative to emotional learning," Wellington, 2009, pp. 651-656.

[17] A. Arami, et al., "A fast model free intelligent controller based on fused emotions: A practical case implementation," Ajaccio-Corsica, 2008, pp. 596-602.

[18] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction Cambridge, MA: MIT Press, 1998.

[19] H. Sagha, et al., "Real-Time IDS using reinforcement learning," Shanghai, 2008, pp. 593-597.

[20] S. B. Shouraki, et al., "Fuzzy Controller design By an Active Learning Method," presented at the 31th Symposium of Intelligent Control, Tokyo, Japan, 1998.

[21] S. Bagheri and N. Honda, "A New Method for Establishing and Saving Fuzzy Membership Functions," presented at the 13th Fuzzy symposium, Toyama, 1997.

[22] S. Bagheri and N. Honda, "Outlines of a Soft Computer For Brain Simulation," presented at the International Conference on Soft Computing Information/Intelligence Systems (IIZUKA'98), Iizuka, Japan,1998.

[23] D. shahmirzadi, "COMPUTATIONAL MODELING OF THE BRAIN LIMBIC SYSTEM AND ITS APPLICATION IN CONTROL ENGINEERING," Texas A&M University, 2005.

[24] H. Sagha, et al., "Reinforcement learning based on active learning method," Shanghai, 2008, pp. 598-602.

[25] P. Hartono and S. Hashimoto, "Active Learning of Neural Network," presented at the Proceedings of 1993 international joint conference on neural networks, Nagoya, Japan, 1993.