BEYOND ADAPTIVE CRITIC - CREATIVE LEARNING FOR INTELLIGENT AUTONOMOUS MOBILE ROBOTS

XIAOQUN LIAO
Center for Robotics Research
University of Cincinnati

ERNEST L. HALL
Center for Robotics Research
University of Cincinnati

ABSTRACT

Intelligent industrial and mobile robots may be considered proven technology in structured environments. Teach programming and supervised learning methods permit solutions to a variety of applications. However, we believe that to extend the operation of these machines to more unstructured environments requires a new learning method. Both unsupervised learning and reinforcement learning are potential candidates for these new tasks. The adaptive critic method has been shown to provide useful approximations or even optimal control policies to non-linear systems. The purpose of this paper is to explore the use of new learning methods that goes beyond the adaptive critic method for unstructured environments. In the adaptive critic family, globalized dual heuristic programming (GDHP) is actually combined heuristic dynamic programming (HDP) and dual heuristic programming (DHP) based on dynamic programming (DP). The objective of this paper is to explore more generalized methods for the adaptive critic family. It is beyond adaptive critic learning theory and defined as creative learning (CL). Creative learning includes all the components in the adaptive critic family, which is to generalize GDHP by modifying the learning rates and utilizing multiple criteria (or critic) and increasing the degree of derivatives of the J (critic) function. A critic element provides only high level grading corrections to a cognition module that controls the action module. In the proposed system the critic's grades are modeled and forecasted, so that an anticipated set of sub-grades are available to the cognition model. The forecasting grades are interpolated and are available on the time scale needed by the action model. The significance of this paper is to better understand the adaptive critic learning theory and move forward to develop more human-intelligence-like components into the intelligent robot controller. Moreover, it should extend to other applications. Eventually, integrating a criteria knowledge database into the action module will develop a real imagination adaptive critic learning module.

1. INTRODUCTION

Intelligence is the most outstanding human characteristic; however, it is still not totally understood and therefore Current researchers are attempting to develop *intelligent robots*. Hall (Hall, 1985) defines an intelligent robot as one that responds to changes to its environment through sensors connected to a controller. The purpose of this paper is to present an idea about new theory of learning called creative learning. This theory is beyond the adaptive controller in that the reinforcement comes from the learning machine rather than from an external critic. Such an approach offers potential solutions to problems in which

the objective criteria is unknown or yet to be discovered. A brief review intelligent robot controller is presented in section 2. Robot learning rules is discussed in section 3. Adaptive critic learning is addressed in section 4 and creative learning theory is described in section 5. Results and conclusions are given in section 6**.**

2. ROBOT NEURAL CONTROLLER

It is the goal of the robot researcher to design a neural learning controller to utilize the available data from the repetition in robot operation. The neural learning controller, based on the recurrent network architecture, has the time-variant feature that once a trajectory is learned, it should learn a second one in a shorter time. In Fig. 1, the time-variant, recurrent network will provide the learning block, or primary controller. The network compares the desired trajectories with continuous paired values for the three-axis robot, at every instant in a sampling period. The new trajectory parameters are then combined with the error signal from the secondary controller (feedback controller) for actuating the robot manipulator arm.

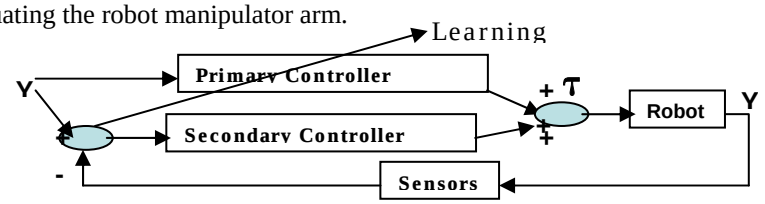

Figure 1. Recurrent neural learning controller

Neural network approaches to robot control are discussed in general by Psaltis *et al* (1988), and Yabuta and Yamada (1992). These approaches can be classified as: (1) Supervised control: A trainable neuromorphic controller reported by Guez and Selinsky (1988) provides an example of a fast, real-time and robust controller. (2) Direct inverse control: is trained for the inverse dynamic of the robot. Kung and Hwang (1989) used two networks on-line in their design of the controller. (3) Neural adaptive control, neural nets combined with adaptive controllers result in greater robustness and the ability to handle nonlinearity. Chen (1990) reported the use of the BP method for a nonlinear self-tuning adaptive controller. (4) Backpropagation of utility involves information flowing backward through time. Werbos's back-propagation through time is an example of such a technique (Werbos, 1990). (5) Adaptive critic method uses a critic evaluating robot performance during training. This is a very complex method that requires more testing (Werbos, 1991). The robot learning rules addressed in the following are applied to all the robot control methods described above.

3. ROBOT LEARNING RULES

3.1 Supervised Learning and Unsupervised Learning

Given a set of input/output patterns, ANNs can learn to classify these patterns by optimizing the weights connecting the nodes (neuron) of the networks. The learning algorithms for weight adaptation can be described as

either supervised or unsupervised learning or reinforcement learning. In supervised learning, the desired output of the neuron is known, perhaps by providing training samples. In unsupervised training, where there are no teaching examples, built-in rules are used for self-modification, in order to adapt the synaptic weights in response to the inputs to extract features from the neuron. Kohonen's self-organizing map is an example of unsupervised learning (Chester, 1993).

3.2 Reinforcement Learning

Sutton, et al (1998) identified four main sub-elements to a reinforcement learning (RL) system: a policy, a reward function, a value function, and, optionally, a model of the environment. And also a summary of RL system is discussed. There are three threads of RL: learning by trial and error , problems of optimal control , and temporal-difference methods. Optimal control problems and its solution using value functions and dynamic programming is also named as adaptive critic learning, which is addressed in next section.

4. ADAPTIVE CRITIC LEARNING

Werbos (1995) summarized recent accomplishments in neurocontrol as a "brain-like" intelligent system. It should contain at least three major general-purpose adaptive components: (1) an Action or Motor system, (2) an "Emotional" or "Evaluation" system or "Critic" and (3) an "Expectations" or "System Identification" component.

"Critic" served as a model or emulator of the external environment or the plant to be controlled, solving optimal control problem over time classified as adaptive critic designs (ACD) (Werbos, 1989). According to modern control theory, dynamic programming is the only exact and efficient method for utility maximization or optimization over future time. In dynamic programming, normally the user provides the function $\mathbf{U}(\underline{\mathbf{X}}(t), \underline{\mathbf{u}}(t))$, an interest rate **r,** and a stochastic model. Then the analyst tries to solve for another function $\mathbf{J}(\underline{\mathbf{X}}(t))$, so as to satisfy some form of Bellman equation, the equation that underlies dynamic programming (Werbos, 2000):

$$J(\underline{X}(t)) = \max_{\underline{u}(t)} (U(\underline{X}(t), \underline{u}(t)) + < J(\underline{X}(t+1)) > /(1+r)) \quad (1)$$

where "<>" denotes expected value.

The nonlinear function approximator **J** is called a "Critic". If the weights **W** are adapted or iteratively solved for, in real time learning or offline iteration, we call the Critic as Adaptive Critic(Werbos, 2000). There are five levels of adaptive critic approach. First, the simplest level is the original Widrow (1973) design. Level one is the Barto-Sutton-Anderson design, which uses a global reward system to train an Action network and "TD" methods to adapt the Critic Level two is called "Action-Dependent Adaptive Critic" (ADAC) (White, et al, 1992). "Brain-like control", represents levels 3 and above. Level 3 is to use heuristic dynamic programming (HDP) to adapt a Critic, and backpropagate through a Model to adapt the Action network. Levels 4 and 5 respectively use more powerful techniques to adapt the Critic – Dual Heuristic Programming (DHP) and Globalized DHP (GDHP) (Werbos, 1995).

HDP and its ACD form have a critic network that estimates the function J (cost-to-go or strategic utility function) in the Bellman equation of dynamic programming, presented as follows (Prokhorov, 1997):

$$J(t) = \sum_{k=0}^{\infty} \gamma^{k} U(t+k) \qquad (2)$$

where γ is a discount factor for finite horizon problems ($0<\gamma<1$), and U(.) is the utility function or local cost. Werbos [HIC] first proposed the idea how to do GDHP. Training the critic network in GDHP utilizes an error measure which is a combination of the error measures of HDP and DHP.

5. CREATIVE LEARNING

5.1 Aadaptive Critic and Creative Learning

The discussion above is a summary of three of the most popular adaptive critic methods for adaptive critic design. Beyond these learning methods (adaptive critic) addressed above, we propose a renovative learning method, called creative learning. Creative learning includes all the components in the adaptive critic family, which is to generalize GDHP by modifying the learning rates and utilizing multiple criteria (or critic) and increasing the degree of derivatives of the J (critic) function. A critic element provides only high level grading corrections to a cognition module that controls the action module. The most important characteristics of creative learning algorithm are: (1) multiple criteria and increased the degree of derivatives of J function, (2) modeled and forecasted critic modules, (3) criteria knowledge database integrated into crittic-action framework. Also, a predictive action module can be realized according to Syam, et al (2002).

5.2 Creative Learning Algorithm

It is assumed that we can use a kinematic model of mobile robot to provide simulated experience to construct a value function in the critic network and to design a kinematic based controller for the action network. A proposed diagram of creative learning algorithm is shown in Fig. 2. In this proposed diagram, there are five important components: criteria (critic) knowledge database, citic network, action network, model-based action and utility funtion. Both critic network and action network can be constructed by using any artificial neural networks with sigmoidal function or radial basis function (RBF). Furthermore, the kinematic model is also used to construct a model-based action in the framework of adaptive critic-action approach. In this algorithm, we build a criteria (critic) database to generalize the critic network and its training process. It's critical especially when the operation of mobile robots is under an unconstured environments. Another component in the diagram is the utility function for a tracking problem (error measurement). In the diagram, Xk, Xkd, Xkd+1 are input and Y is the ouput and J(t), J(t+1) is the critic function at the time. Currently, the simulated result is under this research.

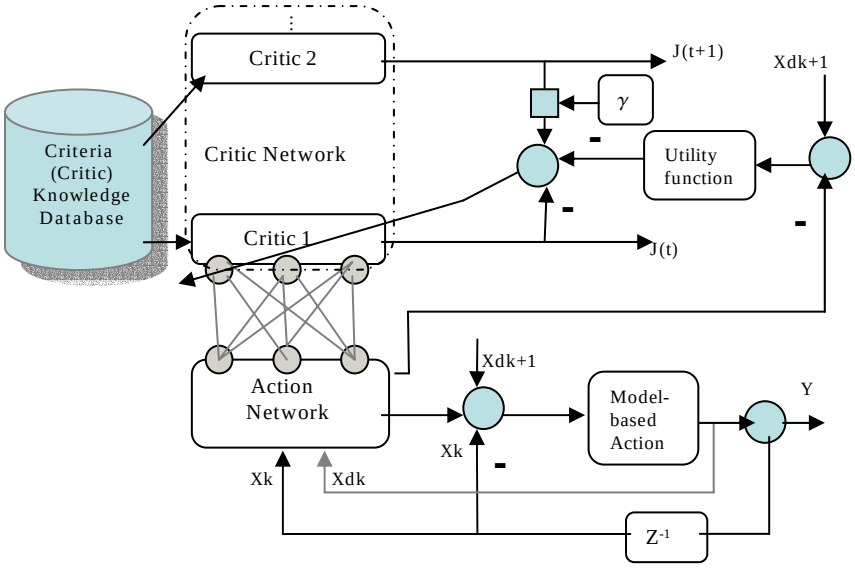

Figure 2 Proposed Creative Learning Algorithms

5.3 Creative Learning Controller

Creative learning is used to explore the unpredictable environment, permit the discovery of unknown problems, ones that are not yet recognized but may be critical to survival or success. It generalizes the highest level of human learning – imagination. As a ANN robot controller, the block diagram of the creative controller can be presented in Fig 3. Experience with the guidance of a mobile robot have motivated this study to progress from simple line following to the more complex navigation and control in an unstructured environment. The purpose in this system is to better understand the adaptive critic learning theory and move forward to develop more human-intelligence-like components into the intelligent robot controller. Moreover, it should extend to other applications. Eventually, integrating a criteria knowledge database into the action module will develop a real imagination adaptive critic learning module.

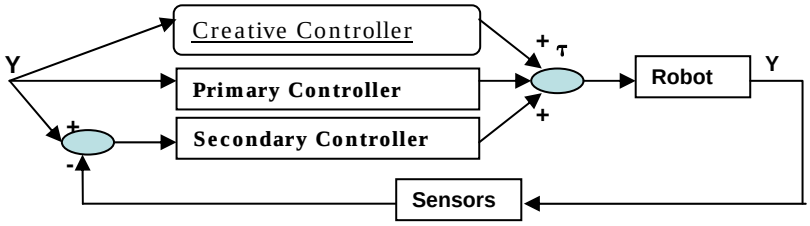

Figure 3 Block diagram of creative controller

6. CONCLUSION

In this paper, a review of learning machine is presented, from robot control strategies to robot learning rules. To design the intelligent robot controller, neural network approaches are addressed above, including supervised control, direct inverse control, neural adaptive control, backpropagation of utility, and adaptive critic method. Beyond the adaptive critic approach, a creative learning theory is proposed in this paper. Creative learning includes all the components in the adaptive critic family, which is to generalize GDHP by modifying the learning rates and utilizing multiple critics and providing model-based action database. The significance of this approach is to generalize the highest level of human learning – imagination. We predict that the creative theory is going to be a real “emotional” or “expectations” component of a “brain-like” intelligent system.

REFERENCES

Chester, M., 1993, Neural Networks: A Tutorial, Prentice Hall, Englewood Cliffs New Jersey.

Chen, F., 1990, "Back-propagation neural networks for nonlinear self-tuning adaptive control", IEEE Control Systems Magazine, pp. 44-48.

Guez, A. and Selinsky, J., 1988, "A trainable Neuromorphic Controller", Journal of Robotic Systems, Vol. 5(4), pp. 363-388.

Hall, E. L., and Hall, B. C., 1985, Robotics: A User-Friendly Introduction, pp. 1-8, Saunders College Publishing, Holt, Rienhart and Wilson, Orlando FL.

Kung, S. and Hwang, J., 1989, "Neural network architectures for robotic applications", IEEE Transactions on Robotics and Automation, vol. 5(5), pp. 641-657.

Prokhorov, D., and Wunsch, D., 1997, “Adaptive critic designs”, IEEE Trans. Neural Networks , Vol. 8, No.5, p.997-1007.

Syam, R., Watanabe, K., Izumi, K., Kiguchi, K., 2002, ``Control of Nonholonomic Mobile Robot by an Adaptive Actor-Critic Method with Simulated Experience Based Value-Functions," Proc. of the 2002 IEEE International Conference on Robotics and Automation, pp. 3960--3965.

Psaltis, D., and Sideris, A., and Yamamura, A. A., 1988, "A multilayered neural network controller", IEEE Control Systems Magazine, vol. 8(2), pp. 17-21.

Sutton, R.S., and Barto, A.G., 1998, Reinforcement Learning: An Introduction, MIT Press, Cambridge, MA, A Bradford Book.

Werbos, P., 1990, "Backpropagation through time: what it does and how it does it", Proceedings of the IEEE, vol. 78, pp. 1550-1560.

Werbos, P., 1991, "An overview of neural networks for control", IEEE Control Systems Magazine, vol. 11(1), pp. 40-42.

Werbos, P., 1995, “Optimal neurocontrol: practical benefits, new results and biological evidence”, Wescon Conference Record, p.580-585.

Werbos, P., 1989, “Backpropagation and neurocontrol: a review and prospectus”, IJCNN Int Jt Conf Neural Network p.209-216.

Werbos, P., 2000, “New directions in ACDs: key to intelligent control an understanding the brain”, Proceedings of the International Joint Conference on Neural Networks v 3 p.61-66.

Widrow, B., Gupta, N., and Maitra, S., 1973, “Punish/reward: learning with a critic in adaptive threshold systems”, IEEE Trans. Systems, Man, Cybemetics, Vol.5 p.455-465.

White, D. and Sofge, D., 1992, Handbook of Intelligent Control, Van Nostrand.

Yabuta, T., and Yamada, T., 1992, "Neural network controller characteristics with regard to adaptive control", IEEE Transactions on System, Man, and Cybernetics", vol. 22(1), pp. 170-176.