Emotion and Learning: Solving Delayed Reinforcement Learning Problem Using Emotionally Reinforced Connectionist Network

Stevo Bozinovski, Liljana Bozinovska*

Electrical Engineering Department, University of Skopje, Macedonia
* Institute of Physiology, Medical Department, University of Skopje, Macedonia
bozinovs@rea.etf.ukim.edu.mk

Abstract

The DRL (Delayed Reinforcement Learning) problem is classical in Reinforcement Learning theory. There were several agent architectures solving that problem including some connectionist architectures. This work describes an early connectionist agent architecture, the CAA architecture, that solved the problem using the concept of emotion it its learning rule. The architecture is compared to another classical DRL problem solving architecture, the Actor/Critic architecture. Possible implication to reinforcement learning theory is pointed out.

Introduction

There are distinguished fields of science, such as Artificial Intelligence, Cognitive Science and Artificial Neural Networks, which were mainly concerned with the *rational aspects* of agents (pattern classification, motion planning, reasoning, learning, etc). In the terminology of agent based system, summarized for example in the work of Wooldridge and Jennings (1994) and in the glossary of intelligent behavior (Mataric 1994) the term "emotion" does not exists.

Recently it becomes evident that the influence of the emotions in the agent cognition, learning and behavior should not be neglected. Emotion has become a concept which should be considered as a necessary component of intelligent behavior that offers a potential for the design of artificial agents. Even more, a question of whether machines can be intelligent without any emotion has been already stated (Minsky 1988).

In this short paper (more details in Bozinovski 1995, Bozinovski et al. 1996) we will describe an agent that actually *implements the concept of emotion in its design.* It is a rather early reinforcement learning system, which uses the emotional concepts of feelings and desirability in its learning rule.

Emotion as internal reinforcement

Reinforcement Learning (e.g. Barto 1997) has become one of the dominant paradigms in learning theory.

A reinforcement learning system which introduced emotional concepts in its learning scheme is the Crossbar Adaptive Array (CAA) agent (Bozinovski 1982). The emotion as a feature was introduced *as a need* to solve the DRL (Delayed Reinforcement Learning) problem, a classical problem in the Reinforcement Learning theory.

The problem is how to learn in an environment which gives reinforcement not after an action, but several steps later; when a reinforcement is received and recognized, the learner could not know to which previously taken action it should be assigned. The DRL problem has many instances, and actually we have a *class of DRL problems.* Learning path in a maze and learning to control to balance a pole are such problems from the DRL class. Both of them were considered within ANW group in 1981, resulting in two different agent architectures to solve them. Figure 1 shows the resulting architectures: the CAA and the Actor/Critic (AC) architecture. It is interesting to analyze them from several viewpoints.

First to note in Figure 1 is that CAA architecture has one memory structure, W, while AC has two memory structures, W1 and W2, of the same size as W. Further, comparing the functional description (Bozinovski 1982 vs. Barto et al. 1983) one can see that CAA uses only one incremental relation, its learning rule

$$w_{aj}(t) = w_{aj}(t\text{-}1) + x_j(t\text{-}1)y_a(t\text{-}1)v(t)$$

where w_{aj} is the connection strength between the situation x_j and the action y_a, and v(t) is the emotional value of the state received in the time step t, as a consequence of the activity $x_j(t\text{-}1)y_a(t\text{-}1)$ in time step t-1. On the other hand, AC uses four incremental relations, two for learning rules and two for memory traces. Also AC has second order learning rules while CAA has only one, first order rule.

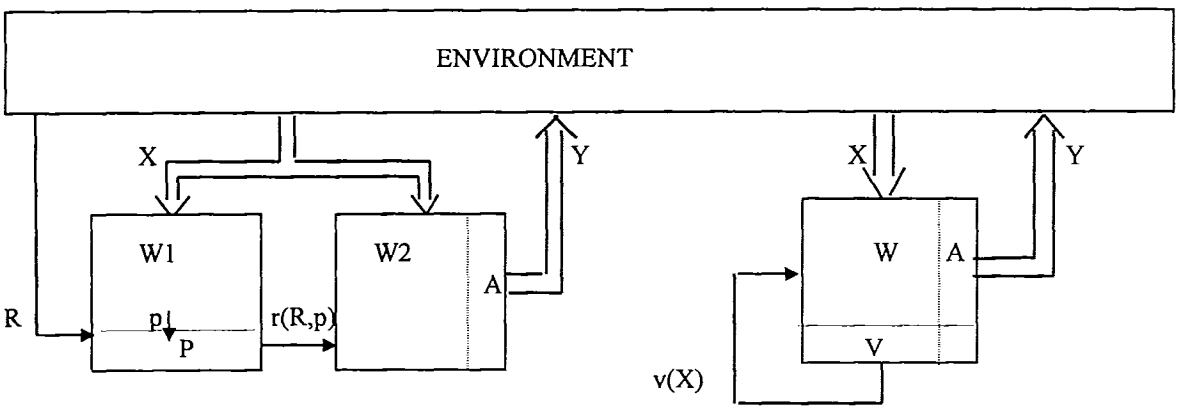

Figure 1.a. Actor/Critic architecture (Barto et al., 1983)

Figure 1.b. Crossbar Adaptive Array architecture (Bozinovski 1982)

Figure 1. A comparison between AC and CAA architecture

Besides the apparent architectural difference between AC and CAA, for the emotion community the most important difference is the philosophy of computing the internal reinforcement. The basic concept for computing the internal reinforcement in AC is *reinforcement prediction*, A separate memory structure, W1, computes the prediction, p, of an *explicit external reinforcement*, R, and supplies an internal reinforcement, r(R,p). In CAA the basic concept is *emotion, desirability* of being in a particular environment situation. CAA does not use the concept external reinforcement; instead it uses the concept of genetic vector, an artificial genome which comes from the genetic environment. That vector defines initial value W(0) of W, which in turn defines *emotional preference* toward the states of the environment. In W(0) an internal state is defined which will produce an emotion if addressed by an environment situation. If that internal state is addressed, a primary reinforcer is defined inside CAA. After that, a *secondary reinforcement* mechanism is taking place: the computed state value v(X) of the current state X is backpropagated to the state from which X is reached, and that state is becoming a new reinforcing state. The importance of the concept of state evaluation used in CAA became evident after the work of Watkins (1989).

Implications to Reinforcement Learning

There are several taxonomies of reinforcement learning systems. The CAA architecture suggests a dichotomial taxonomy by which RL agents can be divided into two classes:

- external reinforcement learning agents, early representative being AC architecture
- emotionally reinforced learning agents, early representative being CAA architecture

References

Barto, A.; Sutton, R.; and Anderson, C. 1983. Neuronlike elements that can solve difficult learning control problems. *IEEE Trans. Systems, Man, and Cybernetics* 13: 834-846.

Barto A. 1997. Reinforcement learning. In O. Omidvar and D. Elliot (Eds.) Neural Systems for Control. p. 7-29, Academic Press

Bozinovski, S. 1982. A self-learning system using secondary reinforcement. In R. Trappl, ed. *Cybernetics and Systems Research,* pp. 397-402, North Holland.

Bozinovski, S. 1995. *Consequence Driven Systems*, Gocmar Press

Bozinovski, S., Stojanov G., Bozinovska, L. 1996 Emotion, embodiment, and consequence driven systems. Proc AAAI Fall Symposium on Embodied Cognition and Action, p. 12-17, AAAI Press

Mataric, M. 1994. Interaction and Intelligent behavior. Ph.D. Thesis, MIT.

Minsky, M. 1988. The Society of Mind, Simon and Shuster

Watkins, C. 1989. Learning from delayed rewards. Ph. D. Thesis, Kings College, Cambridge, England.

Wooldridge, M., and Jennings, N. 1994. Agent theories, architectures, and languages: A survey. In M. Wooldridge, and N. Jennings, Eds. *Intelligent Agents,* Proc. ECAI-94 Workshop, 1-39, Amsterdam, Holland