ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 9, November 2012

Reinforcement learning model for virtual pets

Rodolfo Romero Herrera[1], Leslie Victoria Rodríguez Quiñones[2]

Escuela Superior de Computo del Instituto Politécnico Nacional, Av. Juan De Dios Batiz s/n Col Industrial Vallejo, 07738 México D.F.

ABSTRACT:**Virtual pets are developed with the ability to learn from their environment and interact with the user, which is used for Reinforcement Learning and interaction local or independent. Intelligent agents are used in a virtual environment and receive feedback; we simulate many daily activities of a dog which is essential for the interaction with the user. The use of learning in pets gives a real dimension artificial life and allows the system to evolve intelligently. They give a satisfactory learning in mathematical linguistics considering it as a process. The sequence is repeated at different intervals of time, therefore it is a recursive function, and finally obtained the generalization error, through the concept of distance.**

Keywords:**Learning, computing, artificial intelligence, pet.**

I. INTRODUCTION

Emotions are critical for decision-making, because these influence our daily tasks, such as learning, communication between people [1], but this has been ignored, and that science is not had considered that emotions were an essential part in the scientific and technological development. Sometimes systems were created with frustrating results, partly because the effect has been misunderstood and difficult to measure [2]. Over time various applications have been implemented in affective computing. In this field of study is working on different applications of face recognition, voice, etc.., Where these skills are achieved by a video camera and microphones [3].

One method of learning is used in robots for reinforcement. Before we define the description Agent, and the environment [4].

• Agent in computer is defined as a computer system that lives in a complex and dynamic environment, with the ability to perceive and act autonomously which is able to meet a set of goals or perform certain tasks for which it was designed [5].

• Environment is defined with respect to the agent, as anything that is not him, but that is of interest to carry out a task.

• Reward is a scalar value indicating expected in a situation for the agent, this can get both positive and negative values.

When an agent performs an action, receives a reward value therefore can have consequences later, by partnering with the last action.

Learning theory places special emphasis on the control of behavior by reward or reinforce circumstances, completing their individual prediction process to a wide range of human applications.

From Pavlov and Watson can consider science as a set of attitudes that facilitate observation and experimentation and acceptance of ideas imposed [6].

Moreover Skinner, Hull attacks approaches that try to predict behavior based on internal processes. [7]. As well as that of Watson departs; it accepts thinking and other private behaviors as data sources. However, the method used by Skinner to investigate external variables that control behavior is a causal or functional approach. That is, the laws of behavior involve relationships between cause and effect between the independent variables [environmental facts] and the response variables [dependent]. Therefore we consider that learning is a dependency with the environment.

A key concept of Skinner system is the method of successive approximation, booster responses involved in matching the direction of the final desired response. Shaping depends on response generalization, or tendency of the responses to vary from trial to trial. Each booster can be positive or negative and may be associated with an increased or decreased probability of emission of a particular response.

The concept of Pavlov, Watson and Skinner give us the basis to sustain the system considering it as an agent, its environment and recursive model based on mathematical linguistics.

II. DEVELOPMENT

One of the major problems is the allocation of tasks to a robot or any system either hardware or software is deducted if the result of their actions correspond to what is expected. Therefore to minimize this problem employs the reinforcement learning method, where the nature equation plays an important role in learning justification [9].

This method works with agent-environment interface:
The agent and the environment interact in a sequence of instants of time "t = 0,1,2,3 ..." At each time step t, the

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 9, November 2012

agent receives a representation of the state of the environment S (t) Є S where "S "is the set of possible states.

Based on this, the agent selects an action a (t) ∈ S (t), where S (t), is the set of actions available at time t.

A moment later time, and partly as a result of the action taken, the agent receives a numerical reward, and happens to be in a new state, S (t +1).

To better control the time the agent has a mapping between the representations of states and probabilities to select.

In the methods of the reinforcement learning are described the way in which an agent from its policy gives a result of the experience that get to face the environment. Therefore the aims agent, maximizing long term the sum of the rewards you get. Figure 1 shows the interface-setting agent.

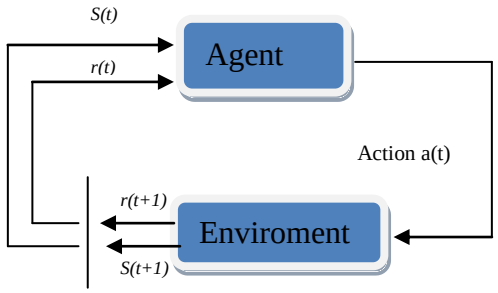

Figure 1.Agente and Environment

Within the reinforcement learning has as main factors the environment and the agent, as these are what make the changes to the system of reinforcement.

The solutions that are obtained depend upon the action of the agent in the environment implies *a (t)*.

Then we have new equation where "*S*" is the solution.

$$s_1 \rightarrow a(t) \qquad (1)$$

S_1 is modified by the environment, generating a new solution, see equation (2):

$$s_2 \rightarrow e\ a(t) \qquad (2)$$

S_1 is modified by the environment, generating a new solution, see equation (2):

$$s_3 \rightarrow s_2 \qquad (4)$$

But if the temperature is obtained:

$$s_3 \rightarrow e^* s_2 \qquad (5)$$

To S_4 solution would be:

$$s_4 \rightarrow e^* s_3 \qquad (6)$$

Where e * represents an environment that is constantly changing. Each new solution depends on how the environment affects the system; so when we get the equation generalize (7).

$$s_G \rightarrow e^* s^* (G-1) \qquad (7)$$

Where *S* * represents a general solution, and G-1 the previous solution.

This equation is called the equation of Nature [8], which we can apply the reinforcement learning because the agent gets a first solution is affected by the environment, and positively or negatively reinforcing learning. That is, the solution becomes learning when S stops changing significantly

III. RESULT

To test the technique used a virtual world for a number of pets on life, where reinforcement is applied by tables and an intelligent agent that interacts with the user.

The graph of Figure 2 shows the value relative for the behavior generated. Where the axis "y" reflects the increased of the behavior learned, and the axis "x" represents the time interval that is generated in this case in minutes.

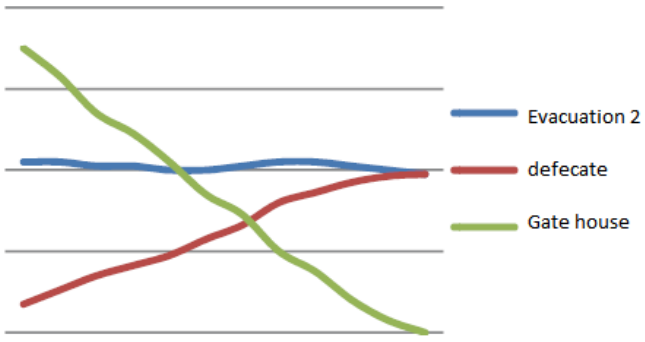

Figure 2. Emotional behavior.

For example in Figure 3 the brown pet dog is distracted with light brown and is therefore difficult to learn this. To achieve that education is necessary to motivate, for

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 9, November 2012

example food or user's affective expressions positively or negatively reinforce their learning

Figure 3. Pet Interaction

In the graph of Figure number 4 we can observe the state of the pet during a time interval. Initially sadness increases due to any need. Later the line was decreased and anger grows predominantly due to dissatisfaction.

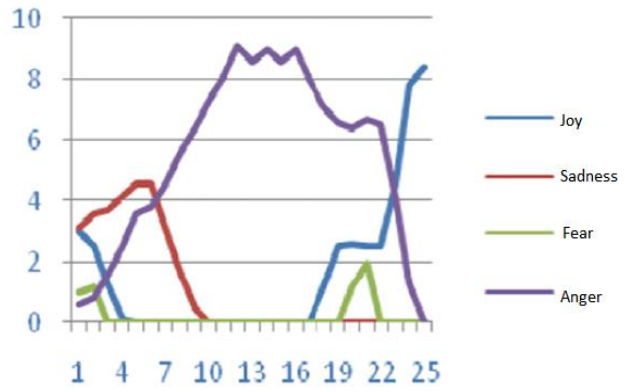

Figure 4. Emotional behavior.

In the reinforcement learning greatly influence emotionson the environment due to the agent, which, as in the case of Figure 4 may be negative, such as fear.

The time in which emotion is generated to produce similar curves Rayleigh or Guassianas so we can approximate them to equations 8 and 9 [9].

$$f(x) = ae^{\frac{(x-b)^2}{2c^2}} \tag{8}$$

Where a, b and c are real constants (a> 0).

The graph of the function is symmetrical bell-shaped. The parameter "a" is the height of the campaign on the point b, where c is the width thereof.

$$f(x|y) = \frac{xexp\left(\frac{-x^2}{2\sigma^2}\right)}{\sigma^2} \tag{9}$$

For example for the "Rage" in the curve of Figure 4 we have for equation 8.

$$f(x) = 9\,e^{\frac{(x-15)^2}{2(25)}} = 9e^{\frac{(x-15)^2}{50}} \tag{10}$$

Where σ^2 represents the variance and x the time.

That is, equations 8 and 9 tell us that emotions have known functions and therefore can simulate different behaviors, and how to help them act in a way, once learning of affective states approximates a bell Gauss.

On the other hand, the generalization error is defined as:

$$E[(g(x) - y)^2] \tag{11}$$

Where:
g (x) is the generalized law
"y" is the royal law
E the mathematical expectation.

But usually not known the generalization error of a model, but we can estimate [10][11].

We made a simulation of reinforcement learning with virtual reality pet interacting with a real environment caused by a user; so it is possible to calculate the generalization error. For thereby creating a concept and after a predetermined time is checked its similarity with the concept learned and then left a time interval until complete twenty samples for this experiment. See Table 1.

Table 1. Generalization Error

Concept	g(x) - y	*Generalization error*
Defecate	2	*0.2*
Go after the ball	2	*0.2*
Eat within an hour	1	*0.05*
Sit	4	*0.2*
Average	2.25	

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 9, November 2012

The generalization error is low so that learning can be acceptable.

Equations 8 and 9 are inserted into 7.

$$S_G \rightarrow e^* \left[a\beta^{-\frac{((x-1)-b)^2}{2c^2}} \right] \quad (12)$$

However, emotions can interact and oppose one another. For this case the "joy" is opposed to sadness and fear can join anger, and these in turn can subtract magnitude to joy. See Figure 4. This situation should lead to interaction of all against all to determine the behavior of a pet. The simulation suggests an interaction such as a butterfly.

Therefore:

$$S_T \rightarrow S_{J-1} - S_s - S_F - S_F \quad (13)$$

Where:

$S_{J-1} =$ Magnitude of Joy
$S_s =$ Magnitude of Sadness
$S_F =$ Magnitud of Anger

IV. CONCLUSION

The reinforcement learning is using intelligent agents that interact with their environment and allow greater realism. Opening a panorama grabbing augmented reality.

Artificial life should be based on methods of nature's own behavior. For this reason the work approaches the equation of nature in reinforcement learning.

Because of this, so that the learning of the pet is compatible with the network of the system variables of behavior, we developed a learning algorithm based on an intelligent agent, and modifying tables reinforcement learning to adapt.

The procedure designed provides realism and instinctive behaviors, which allow for reinforcement learning activities.

The interaction between pets, it is another objective of the system has been performed successfully due to the feasibility of modeling virtual worlds.

The graphs obtained results allow us to approximate to curves Guass and Reyleigh, so that the same can be used to simulate different artificial lives.

The generalization error can be reduced if learning continues for a longer time, repeating the user's activities.

ACKNOWLEDGMENT

This work was partially supported by the National Council for Science and Technology (CONACYT) and the InstitutoPolitecnicoNacional (IPN) of Mexico.

REFERENCES

[1] Isen, A. M., Daubman, K. A., and Nowicki, G. P. Positive affectfacilitatescreativeproblema solving. Journal of Personality and Social Psychology,*Vol.* 52, Issue 6, pp.1122-1131, 1987.
[2] Romero, R., Maquinas que piensan y sienten, Revista digital Universitaria, Vol. 7, Issue 3, 2010.
[3] Pérez, C, Vicente, M.A, Fernández C, Reinoso, O, Gil A., Aplicación de los diferentes espacios de color para detección y seguimiento de caras, Universidad Miguel Hernández, Elche , España, 2010.
[4] Picard, W., Los ordenadores emocionales,Barcelona, Ed.Ariel, 1998.
[5] Brooks, Rodney Allen. Cuerpos y Máquinas. De los Robots a los hombres robots, Barcelona, EdicionesGrupo Zeta, 2003.
[6] Pavlov, I. P., Los reflejos Condicionados,Madrid, Morata; 1997.
[7] Opening Skinner´s., Great psychological experiments of the twentieth century; Printed in the United Stated of America, W.W. Norton & Company, 2005.
[8] Soria, F. , 2010. Una ecuación de la Naturaleza, Available: http://www.fgalindosoria.com/ecuaciondelanaturaleza/una_ecuacion_naturaleza/ec_natu.pdf.
[9] MatousekJiri, JaroslavNesetril; Invitación a la Matematica discreta, Barcelona, Editorial Reverte; 2008.
[10] Hastie, T. Tibshirani, Friedman, J. R. , 2001. The elements of statistical Learning, Springer, Canada, 2001.
[11] González, F. A.,Generalization, Overfitting and Regularization; Departamento de Ingeniería de sistemas e industrial, Universidad Nacional de Colombia, 2007.

Biography

Rodolfo Romero Herrera: Professor - investigator of Mexico in the InstitutoPolitecnicoNacional (IPN). Dr (c) of the School of Mechanical and Electrical Engineering; Director of more than 12 research projects; Author of several paper and books; Research interest in affective computing, mobile computing, artificial intelligence app to electronics and communications.

Leslie Rodriguez Quiñones: Researcher InstitutoPolitecnicoNacional, Graduate of the School of Computer (ESCOM), Participant 4 research projects.