SIMULTANEOUS LEARNING OF PERCEPTIONS AND ACTIONS IN AUTONOMOUS ROBOTS

Pablo Quintía, Roberto Iglesias, Miguel Rodríguez
Department of Electronics and Computing, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

Carlos V. Regueiro
Department of Electronics and Systems, Universidade da Coruña, A Coruña, Spain

Keywords: Reinforcement learning, State representation, Autonomous robots, Fuzzy ART.

Abstract: This paper presents a new learning approach for autonomous robots. Our system will learn simultaneously the perception – the set of states relevant to the task – and the action to execute on each state for the task-robot-environment triad. The objective is to solve two problems that are found when learning new tasks with robots: interpretability of the learning process and number of parameters; and the complex design of the state space. The former was solved using a new reinforcement learning algorithm that tries to maximize the *time before failure* in order to obtain a control policy suitable to the desired behavior. The state representation will be created dynamically, starting with an empty state space and adding new states as the robot finds them, this makes unnecessary the creation of a predefined state representation, which is a tedious task.

1 INTRODUCTION

Robots must be able to adapt its behaviour to changes in the environment if we want them operating in real scenarios, dynamic environments or human's common workplaces. Because of this in this paper we describe a model free learning algorithm, able to adapt the behaviour of the robot to new situations and that not relies on any predefined knowledge. Our system will learn simultaneously how to translate the perceptions of the robot into a finite state space and the actions to perform at each state to achieve a desired behaviour. We are not aware of any other publications with the same objectives.

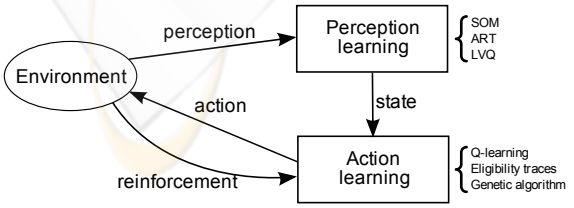


Figure 1: General schema of our proposal with the two modules for perception learning and action learning.

We propose the combination of reinforcement learning (Sutton and Barto, 1998) to learn the actions and a Fuzzy ART network (Carpenter et al., 1991) to learn the states (Fig. 1). Our system will not only learn the actions to execute on each state but also it will learn to classify the situations the robot finds during its operation. Our reinforcement learning based algorithm will be simpler and easier to interpret than other approaches, and the dynamic representation of states will create the state space from an empty set of states. This eliminates the burden of creating an ad-hoc representation for each task. Thanks to this combination of reinforcement learning and Fuzzy ART we will achieve a technique able to learn on-line, adapting the behaviour of the robot to the changes that may occur in the environment or in the robot itself.

2 ACTION LEARNING

Sutton and Barto developed reinforcement learning as a machine learning paradigm that determines how an agent ought to take actions in an environment so as to maximise some notion of long-term reward (Sutton and Barto, 1998).

Reinforcement learning is a very interesting strategy, since all the robot needs for learning a behaviour

Quintía P., Iglesias R., Rodríguez M. and V. Regueiro C. (2010).
SIMULTANEOUS LEARNING OF PERCEPTIONS AND ACTIONS IN AUTONOMOUS ROBOTS.
In *Proceedings of the 7th International Conference on Informatics in Control, Automation and Robotics*, pages 395-398
DOI: 10.5220/0002943203950398

is a reinforcement function which tells the robot how good or bad it has performed, but nothing about the set of actions it should have carried out. Through a stochastically exploration of the environment, the robot must find a control policy – the action to be executed on each state – which maximises the expected total reinforcement it will receive:

$$E[\sum_{t=0}^{\infty}\gamma^t r_t] \quad (1)$$

where r_t is the reinforcement received at time t, and $\gamma \in [0,1]$ is a discount factor which adjusts the relative significance of long-term versus short-term rewards.

Q-learning (Watkins, 1989) is one of the most popular reinforcement learning algorithms, although it might be slow when rewards occur infrequently. What is termed Eligibility Traces (Watkins, 1989) expedite the learning by adding more *memory* into the system. One problem of this algorithms is their dependence of the parameters used, that usually need to be set after a trial an error process.

In this work we present a new learning algorithm based on reinforcement. Our algorithm will provide a prediction of how long the robot will be able to move before it makes a mistake. This raises clear and readable systems where it is easy to detect, for example, when the learning is not evolving properly: basically a high discrepancy between the time before failure predicted and what is actually observed on the real robot. Another advantage of our learning proposal is that it is almost parameterless, so it minimises the adjustments needed when the robot operates in a different environment or performs a different task. The only parameter needed is a learning rate which is not only easy to set, but it is often the same value, regardless of the task to be learnt.

Since we wish to use the experience of each state transition to improve the robot control policy in real time, we shall apply Q-learning, but redefining the utility function of states and actions. Q(s,a) will be the expected time interval before a robot failure when the robot starts moving in s, performs action a, and follows the best possible control policy thereafter:

$$Q(s,a) = E[-e^{(-Tbf(s_0=s,a_0=a)/50T)}], \quad (2)$$

where $Tbf(s_0,a_0)$ represents the expected time interval (in seconds) before the robot does something wrong, when it executes a in s, and then follows the best possible control policy. T is the control period of the robot (expressed in seconds). The term $-e^{-Tbf/50T}$ in Eq. 2 is a continuous function that takes values in the interval $[-1,0]$, and varies smoothly as the expected time before failure increases.

Since $Q(s,a)$ and $Tbf(s,a)$ are not known, we can only refer to their current estimations $Q_t(s,a)$ and $Tbf_t(s,a)$:

$$Tbf_t(s,a) = -50 * T * Ln(-Q_t(s,a)), \quad (3)$$

The definition of $Q(s,a)$, Tbf, and the best possible control policy, determine the relationship between the Q-values corresponding to consecutive states:

$$Tbf_t(s_t,a_t) = \begin{cases} T & \text{if } r_t < 0 \\ T + max_a\{Tbf_t(s_{t+1},a)\} & \text{otherwise} \end{cases} \quad (4)$$

r_t is the reinforcement the robot receives when it executes action a_t in state s_t. If we combine Eq. 3 and Eq. 4, it is true to say:

$$Q_{t+1}(s,a) = \begin{cases} -e^{-1/50} & \text{if } r_t < 0 \\ Q_t(s_t,a_t) + \delta & \text{otherwise} \end{cases} \quad (5)$$

where,

$$\delta = \beta(e^{\frac{-1}{50}} * max_a Q_t(s_{t+1},a) - Q_t(s_t,a_t)). \quad (6)$$

$\beta \in [0,1]$ is a learning rate, and it is the only parameter whose value has to be set by the user.

3 PERCEPTION LEARNING

In reinforcement learning the state space definition is a key factor to achieve good learning times. The state space must be fine enough to distinguish the different situations the robot might find, but at the same time it must have a reduced size to avoid the curse of dimensionality.

The design of the state space is a delicate task, and it is dependent on the problem the robot has to solve. We propose a dynamic creation of the state space as the robot explores the environment (Fig. 1). For this task we have chosen to use a Fuzzy ART artificial neural network (Carpenter et al., 1991). This kind of networks are able to perform an unsupervised online classification of the input patterns without any previous knowledge.

Of the three parametres that are involved in the Fuzzy ART algorithm α, β and ρ – usually called vigilance parameter – the most important is ρ. α and β are almost independent of the task to solve, but the value of ρ will influence the number of states created. If it is too high the Fuzzy ART will create too many classes. If ρ is too low the state representation will be too coarse and the system will suffer from perceptual aliasing, resulting in an increase of the learning time or impossibility to achieve convergence.

Due to space restrictions we can't provide more details of the Fuzzy ART here. Nevertheless further information can be found in (Carpenter et al., 1991).

4 EXPERIMENTAL RESULTS

The two systems showed in Fig. 1 complement each other to find a solution to the learning problem. We will perform several experiments:

a) Evaluate the performance of our learning algorithm described in 2. To do this without the influence of the Fuzzy ART network we used a set of two-layered SOM networks to translate the large number of different situations that the ultrasound sensors may detect, into a finite set of 220 neurones – states (Iglesias et al., 1998).

b) Evaluate the performance of the Fuzzy ART network creating a state space from scratch, using both normalised and not normalised inputs using complement coding (CC) (Carpenter et al., 1991).

We applied our proposal to teach a mobile robot two different tasks: a wall following task; and a door traversal task. The inputs for the Fuzzy ART network will be the inverted readings provided by a laser rangefinder. We reduced the dimensionality to 8 sectors of laser readings 22.5°wide, using the lowest measure as representative of each sector.

The parameters of the learning algorithms used during the learning were: $\beta = 0.288282, \gamma = 0.9, \lambda = 0.869965$. The parameters of the Fuzzy ART were: $\alpha = 0.00001, \beta = 0.0025$.

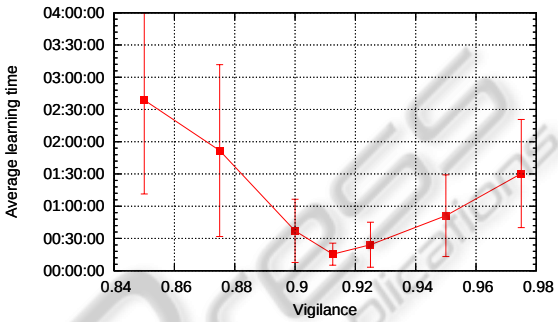
4.1 Wall Following

As said before we will use a static state representation to test the learning algorithms. In order to train the SOM neural networks we used a set of sensor readings collected when the robot was moved close to a wall (Iglesias et al., 1998). For comparison purposes we tested our learning approach against three classical algorithms: Q-learning and two different implementation of eligibility traces: Watkins' Q(λ) (Watkins, 1989) and what is called Naive Q(λ) (Sutton and Barto, 1998). The results obtained after the execution of 15 experiments for each algorithm can be seen in Table 1. The classical learning algorithms performed as expected. Our proposal based on learning the time before failure performed as good as Naive Q(λ). The main advantage of our learning algorithm is to have a more interpretable and simple algorithm, with almost no cost on the learning time.

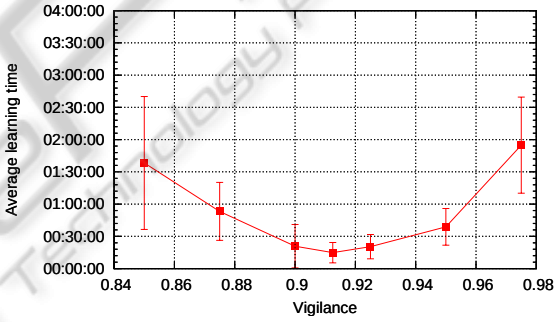
The next step in the experimentation was to combine the Fuzzy ART with the learning algorithm. Considering the previous results, we chose to test the combination of the Naive Q(λ) algorithm and Fuzzy ART. In Fig. 2(a) we can see the variations in the average learning time and std. deviation with different values of the vigilance parameter – ρ . From this

Table 1: Results of the learning of a wall following task with a predefined SOM network (Iglesias et al., 1998).

Algorithm	Learning time	Std. deviation
Q-learning	00:29:37	00:13:59
Watkins's Q(λ)	00:21:35	00:12:32
Naive Q(λ)	00:17:21	00:08:21
Our proposal	00:16:39	00:08:14



(a)



(b)

Figure 2: Results of the wall following task with Naive Q(λ) and Fuzzy ART network without (a) and with (b) complement coding.

results we can extract that the valid range – learning times lower than 1 hour – for the vigilance is approximately [0.900, 0.950] and that the best values are around 0.9125. Fig. 2(b) shows the results of the experiments if the inputs of the Fuzzy ART are codified in complement coding. We can see that the use of complement coding does not reduce significantly the learning time achieved by the optimal vigilance value, but it does improve the learning times if the vigilance parameter is not the optimal.

The best value for the vigilance parameter found in this experiments – 0.9125 – can serve as a good starting point for the use of the Fuzzy ART in other tasks. This value will be used to learn other tasks.

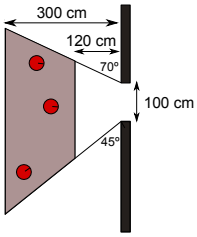


Figure 3: Experimental scenario for the door traversal task. The initial positions of the robot were within the shaded area.

Table 2: Results of the learning of a door traversal task with Naive Q(λ) and Fuzzy ART.

Vigilance	Beams	CC	Average learning time	Average deviation	Average states
SOM	8		01:36:53	00:57:55	221
0.9125	8	No	NA	NA	63.22
		Yes	01:37:12	01:00:55	94.87
	181	No	01:57:48	01:05:31	134.33
		Yes	00:39:56	00:23:20	82.87

4.2 Door Traversal

The door traversal task (Nehmzow et al., 2006) was learnt in the experimental scenario shown in Fig. 3. First we tested the system using the same SOM neural network that was used to learn the wall following task. The robot achieves a good control policy after a learning time of 01:36:53 with high variability – 00:57:55. Using our system the results are equivalent if complement coded is used, without complement coding the system is unable to learn in reasonable time (Table 2).

With the first experiments we found out that if we use the same input as in the wall following task the door was not visible from several positions. To have a better perception of the door we decided to use all 181 laser readings. This improves the times significantly, the average learning time is reduced to 00:39:56 and the std. deviation lowers to 00:23:20. As can be seen in Table 2, the dynamic representation scales very well with the increase in the dimensionality. Complement coding is the appropriate choice for the Fuzzy ART inputs.

5 CONCLUSIONS

Through reinforcement learning the robot is able to learn on its own – through trial an error interactions with the environment – using only the feedback provided by a very simple reinforcement function. The learning algorithm developed in this paper represents a simpler and more interpretable solution to the learning problem. The algorithm requires less parameters and its meaning is more straightforward – the expected time before the robot commits an error.

But one of the main problems of applying reinforcement learning in robotics is the state space definition. In this paper we showed how we can use a Fuzzy ART neural network to dynamically create the state space while the reinforcement learning algorithm learns the actions to execute on each state.

The use of a dynamic representation of states does not suppose an increase in the learning time, in fact it reduces the learning time in comparison to the use of a predefined and static state representation if a good vigilance value is chosen. We also proved that this dynamic state representation scales well with size of inputs. But the main advantage of this approach is that there is no need to create an ad-hoc state representation for the task. Creating a predefined state representation requires gathering a training and test data set, training the network and validating the network. This must be repeated until we obtain a good network for our purpose.

Our proposal was used to solve two different and common tasks in mobile robotics: wall following and door traversal. The experimental results confirm that our proposal is valid.

ACKNOWLEDGEMENTS

This work has been funded by the research grants TIN2009-07737, INCITE08PXIB262202PR, and TIN2008-04008/TSI.

REFERENCES

Carpenter, G. A., Grossberg, S., and Rosen, D. B. (1991). Fuzzy art: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, 4(6):759–771.

Iglesias, R., Regueiro, C. V., Correa, J., Sánchez, E., and Barro, S. (1998). Improving wall following behaviour in a mobile robot using reinforcement learning. In *ICSC International symposium on engineering of intelligent systems (EIS'98)*, Tenerife (España).

Nehmzow, U., Iglesias, R., Kyriacou, T., and Billings, S. (2006). Robot learning through task identification. *Robotics and Autonomous Systems*, 54:766–778.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.

Watkins, C. (1989). *Learning from Delayed Rewards*. PhD thesis, University of Cambridge, England.