ESANN'2006 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2006, d-side publi., ISBN 2-930307-06-4.

Some Experimental Results with a Two Level Memory Management System in the Multilevel Darwinist Brain

F. Bellas, J.A. Becerra and R.J. Duro *

Grupo Integrado de Ingeniería - Universidade da Coruña
Ferrol - Spain

Abstract. This paper provides a description and discussion of several experiments carried out with simulated and real agents that operated with the Multilevel Darwinist Brain cognitive mechanism including a two level memory management system. The agents interacted with real environments, including teachers, and the results show the interplay between the parameters that regulate replacement strategies in both, short term and long term memories. This type of structures allow the agents to learn autonomously, paying attention to the relevant information and to transform data into knowledge, creating subjective internal representations that can be easily reused or modified to adapt them to new situations.

1 Introduction

With the aim of increasing generality and flexibility with respect to that of traditional learning architectures for autonomous agents, especially in regards to their value systems [1] [2], and in the line of cognitive developmental robotics [3] we have developed a Darwinist based cognitive architecture called the Multilevel Darwinist Brain (MDB) [4]. It allows a general autonomous agent to decide the actions it must apply in its environment in order to fulfill its motivations.

This development has been further enhanced by the addition of a two level memory system, whose justification and operation was described in the first part of this two paper series. This second part reports several experiments carried out on simulations and on real robots operating in real environments. The objective is to test the properties of a memory system made up by a short term memory (STM), a long term memory (LTM) and an interaction mechanism that through the analysis of instabilities in predictions is able to regulate their operation.

Thus, the architecture should allow the agent to autonomously extract the relevant information for the creation of all the models involved in its cognitive architecture, including its current value system through a satisfaction model, and discard the rest. In addition, the agent must be able to transform data into knowledge creating subjective internal representations that are usable and can be accessed in the future. This means that the acquired knowledge must be used to adapt to repeated situations or to facilitate learning processes in new situations, including the induction of models extracting conclusions from guided behaviors.

[*] Work partially funded by MEC of Spain under project VEM2003-20088-C04-01 and Xunta de Galicia through project PGIDIT03DPI099E.

2 Experiments

What follows describes a set of experiments that provide an idea of how the MDB with the memory structure works; what is relevant and what it can do. These experiments are divided into two parts. The first part deals with the MDB using only the STM. That is, the most basic sensorial data recollection memory. In it we try to show the need for the proposed replacement strategy and the usefulness of basing it on saliency and temporal relevance parameters in terms of obtaining the most general models possible from the sparse sampling an agent usually has of its environment. Special attention is paid to the fact that the data in the short term memory will simultaneously induce different satisfaction models depending on the sensor sets used. This effect will allow for an agent to continue with a job it has been trained to do even without the teacher stimulus.

In a second block we introduce a long term memory and test its adequateness for the recovery of pre-learnt models as well as the repercussion of the interaction between short and long term memories on the global performance of the agents. This interaction occurs when the long term memory management system detects instabilities in predictions of the models or increases in model access to LTM and acts over parameters that regulate the operation of the STM making thee model generating data more or less general.

In the case of the long term memory, it stores models that have proven to be good predictors for extended periods of time and their associated context (STM contents). Every time a model is a candidate for the LTM, it is phenotypically compared to all the other models present in LTM by executing crossed predictions with the others' contexts. If predictions are different enough, it is assumed that the models correspond to different situations and are thus introduced in LTM.

2.1 MDB using only STM

We have prepared an experiment that aims to test the data acquisition and selection capabilities on a real robot in a real environment interacting with humans. The idea is to introduce the fact that more than one model is simultaneously being generated. That is, there are world models for different sensorial information and satisfaction models depending on different relationships between sensorial data and real feedback.

This experiment was carried out using a Pioneer 2 robot. This is a wheeled robot with a sonar array around its body and with a platform on the top in which we placed a laptop where the MDB was executed. The experiment consists on a teacher that provides commands to the robot in order to capture an object. The seven possible commands were translated into the seven musical notes perceived by a microphone. Initially, the robot has no idea of what each command means. After sensing the command, the robot acts and, depending on the degree of obedience, the teacher provides a reward or a punishment

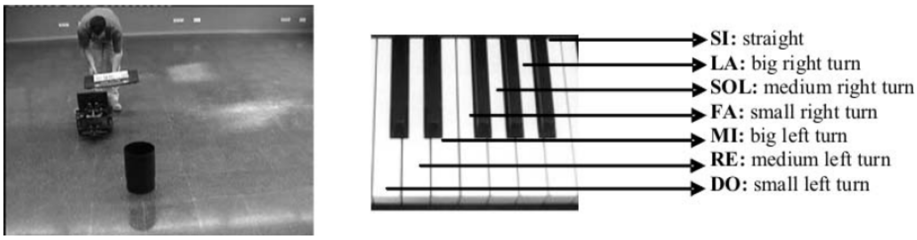

Fig. 1: Experimental setup (left) and translation of commands into musical notes (right)

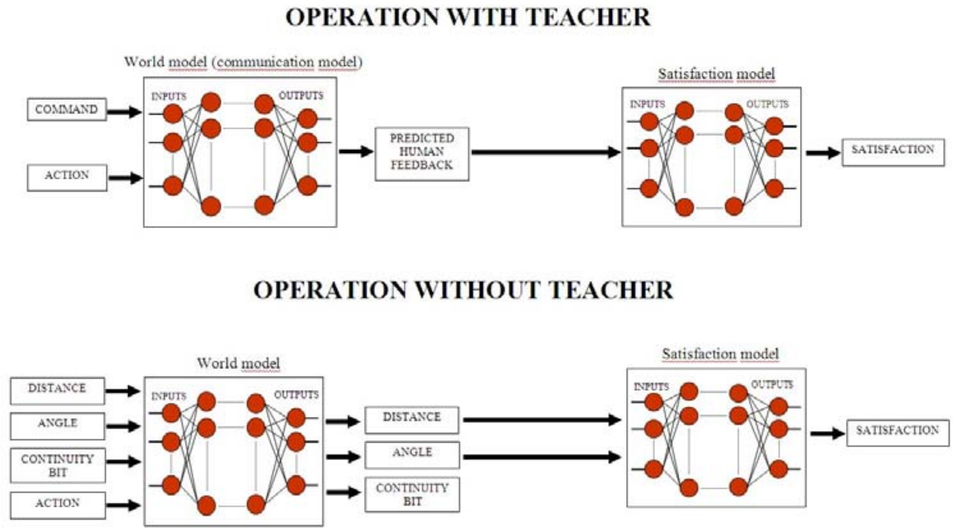

Fig. 2: Models for operation with (top) and without (bottom) teacher

through a numerical value as a pain or pleasure signal introduced via keyboard. This experimental setup is shown in the left image of Figure 1.

The main objective was to show that the information stored in the STM (action-perception pairs) through the replacement strategy allows the agent to create, at least, two kinds of models that come about when modeling different sets of sensors: one related to the sound sensor for the operation when the teacher is present and an induced model or models relating to the remaining sensors. The robot will have to resort to these models when the teacher is not present in order to fulfill its motivations. It is important to note, that when talking about models we are including both, world/internal models and satisfaction models. The latter are the ones that provide the value system for the action selection mechanism to select what is better. The desired induced behavior appears from the fact that each time the robot applies the correct action according to the teacher's commands, the distance to the object decreases. This way, once the teacher disappears, the robot can continue with the task because it developed a satisfaction model related to the remaining sensors that tells it to perform actions that reduce the distance to the object.

Figure 2 displays a schematic view of the two world model and satisfaction model combinations that arise in this experiment. The top models are related with the sound sensors and the bottom models with the sonar ring. Consequently, the top world model of Fig. 2 is related with robot-teacher communication, so we will call it *communication model*. It has 2 inputs (Fig. 2 top) and 1 output:

Commands provided by the teacher: that have been translated into musical notes using the encoding shown in in Figure 1 (right). This encoding is not pre-established and we want the teacher to make use of any correspondence it wants.

Action applied: after perceiving a command, the robot can apply one of the actions in figure 2.

Predicted human feedback: depending on the degree of fulfillment of a command, the distance covered and the angle to the object (we want the robot to capture the object

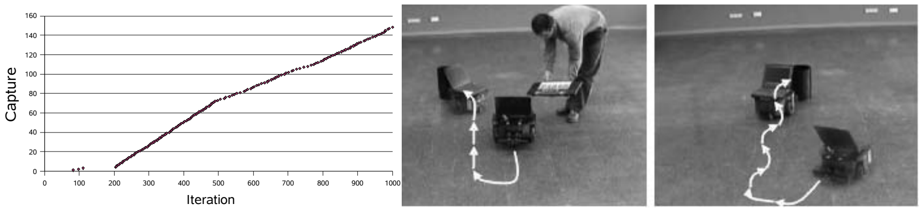

Fig. 3: Number of object captures through iterations (left graph) and real operation of the Pioneer 2 with and without teacher (middle and right images)

frontally), the teacher must reward or punish the robot. To do this, a numerical value (from 0 to 11) introduced through a keyboard is used as a pain or pleasure signal.

In this first case when the teacher is present, the satisfaction model is trivial because the satisfaction coincides with the output of the *communication model*, this is, the reward or punishment. In the case of the sonar ring, for the sake of simplicity, the sensors have been virtualized into 3 values. Therefore, the model (Fig. 2 bottom) has 4 inputs and 3 outputs:

Distance: from the robot to the object, provided by the sonar array.

Angle: with respect to the motion axis of the robot

Continuity bit: virtual sensor necessary because of the symmetry of the sensed angles in this kind of circular robots

Action applied: same as in the communication model.

The 3 outputs are de predicted distance, angle and continuity bit after applying the input action.

In this second stage the satisfaction model is more complex and provides the satisfaction value from the distance and angle, which had a direct relation with the rewards or punishments in the learning stage.

In this particular experiment, the four models are represented by multilayer perceptron ANNs that were adjusted using the PBGA genetic algorithm [5]. So, in this case, the MDB executes four evolutionary processes over four different populations of models each iteration. The main results from this experiment are shown on Figure 3 (left) which displays number of object captures (distance to the object less than 10 cm and angle in the interval -10º to 10º from the frontal part of the robot to the object) as the robot interacts with the world, taking into account that the teacher provides commands until iteration 500. It is clear from the figure that after a few unsuccessful interactions while the information in the short term memory is insufficient, the models become adequate and the object captures become continuous. In addition, when the teacher disappears the robot continues capturing the object in the same way the teacher had taught it. Consequently, we can say that the induced learning of the models has been successful. The decrease in the slope of captures implies that the robot takes a larger number of iterations (movements) to reach the object using the induced models. This is because it has learnt to decrease its distance to the object and not the fastest way to do it (these models aren't based on obedience but on distance increments).
Figure 3 (middle and right images) displays a real execution of actions. In the left part, the robot is following commands by a teacher; in the right it is performing the behavior without any commands, just using its induced models. It can be clearly seen that the behavior is basically the same although a little less efficient.

2.2 MDB using STM and LTM

We have developed a simulated experiment based on the previous one where a model of the Pioneer 2 robot must reach an object. The world and satisfaction model types are the same as those in Figure 2 (bottom). In this case we have included the LTM and its replacement strategy, so we are using a complete version of the MDB. The main objective of the experiment was to test the behaviour of the MDB with and without using the acquired knowledge represented by the models stored in the LTM. Figure 4 displays the evolution of the mean squared error in the prediction of the angle made by the best world model over all the samples stored in the STM each iteration. The pointed line represents an execution of the MDB where the models stored in the LTM are injected as seeds in the population of world models while the continuous line represents an execution where the models were not injected. In addition, every 600 iterations we simulate a failure in the robot (interchanging distance and angle inputs and decreasing their sensed value by 80%) in order to provoke a change in the world models.

As shown in Figure 4, when using the LTM, the MSE obtained for the angle in all the zones with the failure (iterations 600-1200, 1800-2400, 3000-3600 and 4200-4800) is very low. This happens because the system reaches this error level between iterations 600 to 1200, and the world model that provides this error is stored in the LTM by the replacement mechanism according to its relevance and accuracy. From this point forward, this model is injected in the population of the world model's evolutionary algorithm and, consequently, the same error level is immediately reached the next times. The execution that doesn't use LTM (continuous line) needs to relearn from the initial population each time. For example, in the interval 3000-3600 this execution reaches the same error level as the other one but as the model is not stored, the next time in the same zone (4200-4800), the error level is higher again. As shown in Figure 4, there are error peaks every 500 iterations when the world changes. The memory management system detects this error increase and changes the replacement strategy of the STM to a FIFO type strategy ($K_d= K_r= K_c=0, K_t=1$) that purges the old samples. Thus, the STM can rapidly purge the action perception pairs it has from the previous situation, that do not represent the current reality, and make room for samples corresponding to the new reality.

To complement these simulated results, an experiment was also carried out using a real robot. The experiment is the same as presented in section 2.1 with the Pioneer 2 robot, but now, once the robot has learnt to follow the commands to reach the object (first 500 iterations), the teacher changes the task and now it leads the robot to avoid the object (from iteration 500 to 1000). This implies that the robot must learn a new satisfaction model. In

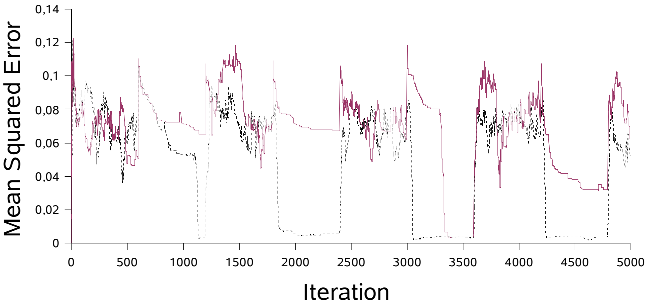

Fig. 4: Evolution of the mean squared error in the prediction of the angle over the STM each iteration, with (pointed) and without (solid) LTM.

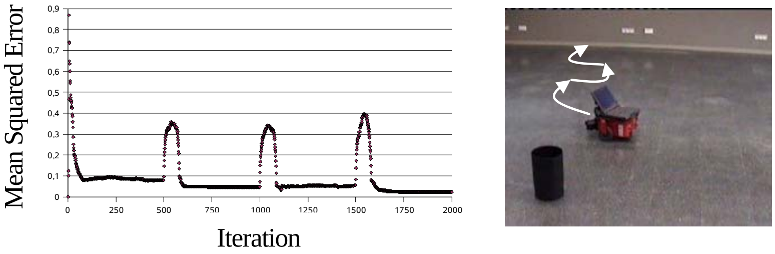

Fig. 5: MSE with a change of satisfaction every 500 iterations (left). The right image represents the robot operation when the satisfaction guides it to escape

iteration 1001 the teacher starts again with the first behavior and guides the robot to the object again. Figure 5 shows the evolution of the MSE for the satisfaction model with changes in the satisfaction every 500 iterations. As expected, the adaptation of the models is very fast after changes, and once an error level is acquired it is never lost (the use of knowledge is correct). The error peaks every 500 iterations clearly mark the changes in the satisfaction model, but the memory management systems works as expected (purging the STM when a change is detected) so there is no pollution in the STM and thus independent models are obtained automatically for each zone.

3 Conclusions

As a follow up of a previous paper in which the foundations of the operation of the MDB with a two level memory structure were introduced, this paper has presented some real experimental results of the application of such a mechanism. The results show the importance of the interplay between memories in an on-line learning system. The data stored in the Short Term Memory are modeled and stored in the LTM as knowledge the agent can use in future learning situations. The replacement mechanisms of these memories are very interdependent because the acquisition of knowledge depends to a large extent on the relevance of the data. The results also show that in real practice, the agents learn in an autonomous manner and pay attention to the relevant information of each situation. In addition, the agent creates subjective internal representations that allow it to operate with and without teachers making use of the available sensorial information and the different models induced during learning.

References

[1] J. Weng, Developmental Robotics: Theory and Experiments, *International Journal of Humanoid Robotics, vol. 1, no. 2,* 199–236, (2004)

[2] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur and E. Thelen, Autonomous Mental Development by Robots and Animals, *Science, vol. 291*, no. 5504, pp. 599 - 600, Jan. 26, 2000.

[3] Asada, M., MacDorman, K. F., Ishiguro, H., Juniyoshi, Y., Cognitive Developmental Robotics as a New Paradigm for the Design of Humanoid Robots. *Robotics and Auton. Systems, V. 37,* pp. 185-193 (2001).

[4] F. Bellas, A. Lamas, R.J. Duro, Multilevel Darwinist Brain and Autonomously Learning to Walk, *Proceedings CIRAS2001*, pp 392-398, (2001)

[5] F. Bellas, R. J. Duro,. Statistically neutral promoter based GA for evolution with dynamic fitness functions. *Proceedings of IASTED2002,* pp 335-340 (2002)