Reinforcement Learning in MirrorBot

Cornelius Weber, David Muse, Mark Elshaw, and Stefan Wermter

Hybrid Intelligent Systems, SCAT, University of Sunderland, UK
{cornelius.weber, david.muse, mark.elshaw,
stefan.wermter}@sunderland.ac.uk
http://www.his.sunderland.ac.uk

Abstract. For this special session of EU projects in the area of NeuroIT, we will review the progress of the MirrorBot project with special emphasis on its relation to reinforcement learning and future perspectives. Models inspired by mirror neurons in the cortex, while enabling a system to understand its actions, also help in the solving of the curse of dimensionality problem of reinforcement learning. Reinforcement learning, which is primarily linked to the basal ganglia, is a powerful method to teach an agent such as a robot a goal-directed action strategy. Its limitation is mainly that the perceived situation has to be mapped to a state space, which grows exponentially with input dimensionality. Cortex-inspired computation can alleviate this problem by pre-processing sensory information and supplying motor primitives that can act as modules for a superordinate reinforcement learning scheme.

1 Introduction

Brain-inspired computation has the prospect of unprecedented control of artificial agents in addition to giving insights into neural processing. In the MirrorBot project, cortical mirror neurons which link perception and action have been chosen as a topic of study and a source of inspiration for building artificial systems. Mirror neurons which have been found in the motor cortex of the monkey are not only active when a monkey performs an action, but also when it observes the corresponding action being performed by somebody else (e.g. [1]). Thus, they have sensory neuron properties. This justifies that we generalise models of the sensory cortex, in particular from vision, to the motor cortex. Such models can learn, instead of a representation of a visual image, a representation of a sensory-motor mapping that has been given as input during learning and that may originate from a reinforcement learner. In reinforcement learning, the input state space grows exponentially if actions are extended. Here a motor cortex module can become a replacement. The hierarchical structure of the cortex furthermore suggests a capability of action organisation, and if given motivational input such as the reward (or Q-value) used for reinforcement learning [2] then the cortex might represent and act in response to such values. A view emerges that a reward-driven reinforcement module is surrounded by a cortex that not only pre-processes its input but also learns to understand, duplicate and anticipate it. We use a simple, but expandable robot docking manoeuvre as an example of a real world demonstration.

2   A Visually Guided Robotic Docking Task

Grasping of an object is a fundamental task for monkeys and humans. The robot equivalent is the docking, where it has to approach a table in a fashion that it can grasp an object lying on it. Figure 1 shows the geometry. A video can be seen at: `http://www.his.sunderland.ac.uk/robotimages/Cap0001.mpg`.

We have managed to control the robot performing this task based almost entirely on neural networks. Figure 2 shows the model. This consists of several modules which have partially been trained independent of each other.

First, the lower visual system consists of the mapping from the raw pixel image I to the hidden feature representation u. This is trained unsupervised according to a generative model. Accordingly, the image has to be generated from the hidden code via feedback weights which are used only during training.

Second, the location associator weights from u to the area representing the perceived location p are trained supervised with the target object position given during training. After training the location can be filled in if it is missing, thus the attractor network does pattern-completion to localise the object.

Third, an action strategy is learnt by reinforcement using an actor-critic paradigm. Its input is the state f which is constructed as the outer product of p and a vector representing the robot angle φ w.r.t. the table. The critic value c represents the goodness of the current robotic state which is rewarded if the target object is at a graspable position, i.e. perceived near and in middle of view, and while $\varphi = 0$. During learning, the motor actions acquire a strategy to reach this goal [3]. In the following we will show that an alternative module can copy this action strategy to replace the reinforcement learner after task acquisition.
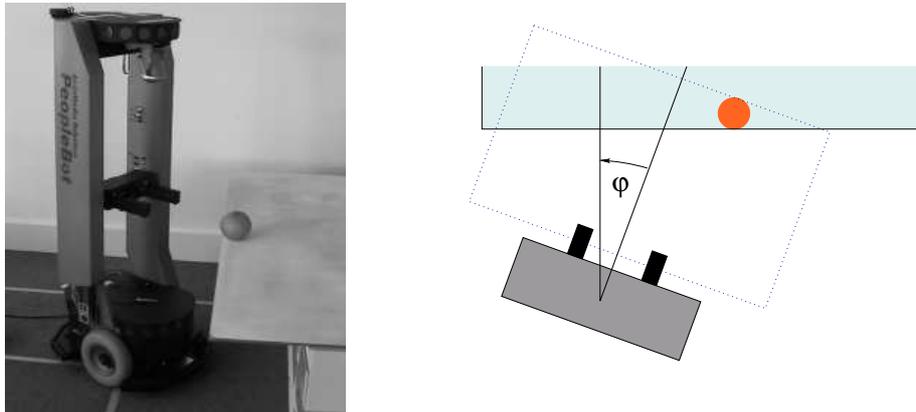
Fig. 1. Left, the PeopleBot robot aiming to grasp an orange fruit from a narrow table. Its camera is below the top-plate and is assumed fixed throughout learning and performance. Right, the geometry of the scenario with the table and target, above, and the robot with its grippers, below. The robot's input is the visual field (outlined by a dotted rectangle) and its angle φ to the table, obtained from internal odometry.
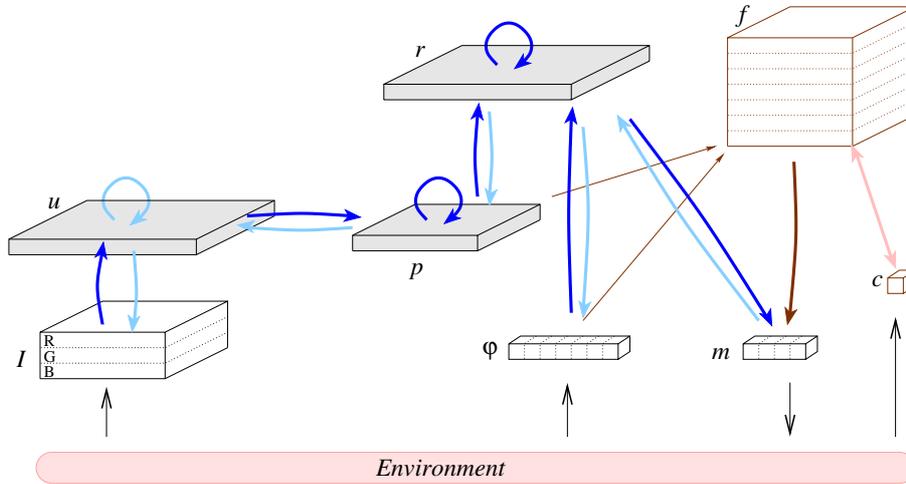
Fig. 2. The neural network which performs visually guided docking. The information flow is as follows: An RGB colour image I from the robot camera is given as input. A representation u is obtained on what we would identify as a V1 visual area, and from this we obtain the perceived location p of the target object within the image. This location and the rotation angle φ of the robot, together contain the information which is collated as state space vector f. This is evaluated during learning by the critic c, and the motor actions m drive the robot's wheels. After task acquisition, the motor cortex representation r binds sensory-motor associations, and can produce actions m based on inputs p and φ by itself. This makes the state space available for learning other tasks. Thick arrows represent trained weights, the lighter of which are used only during learning. Shaded areas are supposed to belong to the cortex.

3   Motor Cortical Neurons Performing Docking

A further fourth module is a cortical representation r which associates this area's inputs p, φ and m. The combined input allows it to perform the stimulus-response mapping already performed via the state space. The intra-area attractor network connections are trained for prediction, allowing the network in addition to perform mental simulation. The idea is that automatic performance of the motor primitive by the motor cortex module makes the state space redundant and thereby makes it available to learn other tasks. A video showing the robot controlled by the simulated motor cortex can be seen at: `http://www.his.sunderland.ac.uk/supplements/NN04/MOV01065.MPG`

   We propose to identify the model's state space with the basal ganglia, as these have been linked with reinforcement learning. There is biological evidence that the basal ganglia are active only during early phases of task acquisition [4].

4   Mirror Neurons for Multiple Actions

One advantage of the cortical modular motor action over reinforcement-trained agents is that cortical representations can easily be structured hierarchically, allowing multiple and composed actions to be represented. We produced three simulated robotic actions, "pick" which corresponds to the docking, "lift" during which (after picking an object) the robot would move back and turn and "go" during which the robot avoids objects and wanders around. We designed the environment and the robotic perception so that all behaviours act based on the same sensory input. The robot rotation angle φ and the distance to the wall and object p are contained in a "high-level vision" sensory input array.

Figure 3 shows the network architecture which performs these tasks which we have implemented on a simulator [5]. A top level area is added containing a vector s, implemented as a SOM [6], which binds language input l together with a representation r that performs previously acquired sensory-motor bindings. Given language as input and thereby influencing the winner among s will influence the sensory-motor mapping. Vice versa, if complete sensory-motor stimuli are given, then the winner among s identifies the action which is being performed. Production and recognition share the same neural substrate as is the case for mirror neurons [1]. Also, action words are topographically arranged [7].

5   Extension of Docking via a Long-Range Strategy

The visually guided docking described in Section 2 requires that the robot is very close to the table and that the target object is visible. We are currently implementing a long-range table approach by reinforcement learning. While the robot's camera cannot find and identify the target object from a large distance, the table can be identified. This is particularly easy by an additional omni-directional camera fitted on top of the PeopleBot robot. Several design implementations are considered to integrate long-range and short-range docking.

(i) A straightforward approach is a monolithic state space spanning long- and short-range sensory input. In this case, however, the state space would become too large. In order to overcome these problems, it has been proposed to use adaptive state recruitment schemes [8][9].

(ii) A second approach is to train long-range and short-range behaviours separately using separate modules and to combine them sequentially. While this specific partitioning scheme might sound arbitrary, the switch between the use of the omni-directional camera for the long-range and the standard robot camera for the short-range clearly marks a boundary for the behaviour change. In humans, the switch might not be determined by the use of different sensors, but of actuators instead, such as the use of legs for long-range approach and of arms for the short range. Having defined behaviour modules, or *partial policies* which accomplish subtasks, it has been proposed to hierarchically implement a superordinate reinforcement scheme that acquires a policy for optimally switching between the subtasks as if they were primitive actions [10][11], see also [12]. This scheme, however, seems too powerful if there are just two modules.
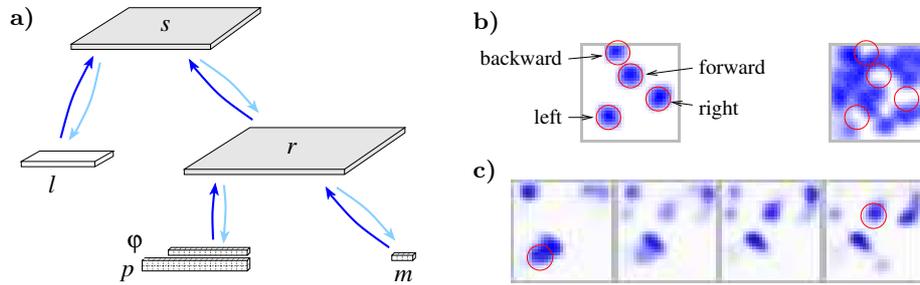
Fig. 3. a) The model architecture for multiple actions. The sensory-motor area representing r and binding sensory inputs (p, φ) with motor actions m stems from the model in Figure 2. New is the top-level SOM area where s associates language input l with a certain motor program r. **b)** Left, the sensory-motor area and its innervation from the four motor area units. Each motor unit projects only onto a small (circled) region on this area. Right, the sensory-motor area and its combined innervation from the sensory inputs. One can see that the four areas which receive motor input (circled) are avoided by sensory input. **c)** Receptive fields of four SOM area units in the sensory-motor area. It can be seen that each SOM unit receives innervation from patchy regions in the sensory-motor area. The leftmost unit contains a sub-region (circled) that also receives input from the "left" motor unit, while the rightmost unit has a coinciding region (circled) with the "forward" motor unit. Thus the SOM area units perform dynamical feature binding, associating slightly different sensory input with different motor actions. The four units shown are all active during the "go" action; SOM units corresponding to different actions perform different bindings.

(iii) Contributing to such a hierarchical implementation, a behaviour module, or *action sequence* may be represented on the motor cortex, as we have proposed in Section 3. In this case the motor cortical units coding for that action sequence would be addressable by the reinforcement module just as single motor units are in the canonical implementation. The state space then would not need to consider any input that is accounted for by the cortical units. Both, long-range and short-range docking could be implemented by such motor units.

6   Discussion

Mirror neurons may play a major role in a distributed language representation of actions [7] by their multimodality [1]. In the MirrorBot project, also a modular neural architecture was devised to parse and understand a sentence [13] which we will integrate with the model described here. Further improvements have been done on attractor network models for visually focussing objects [14].

We have seen in Section 2 that a visual cortex-inspired module can deliver an object representation as required as input to a state space. This requires a fixed camera so that a visually perceived position can readily be used in the robot's

motor coordinates. We are currently developing a coordinate transformation attractor network which will allow the camera to be rotated while the robot is docking. It associates *(i)* the visually perceived object location and *(ii)* the camera pan-tilt angle with *(iii)* the body-centred position of the target object. This is another strategy to extend the action range and limiting the state space.

In Section 3 we have seen that a motor-cortex inspired module can obviate the reinforcement module, and Section 4 demonstrated the potential of cortical action organisation. Our cortical models are inspired by the theory of generative models which reconstruct training data. Therefore, a "teacher" module, such as the reinforcement trained module, is required for any new action that the cortex then performs habitually. If the cortex receives the motivational, reward value used for reinforcement learning as additional input, then it is able to specifically perform such state-action associations which lead to a high reward [2]. With its associative and predictive capabilities, the cortex might directly use incoming stimuli to predict motivations of the agent, and enable a teleological behaviour.

References

1. Rizzolatti, G., Fogassi, L., Gallese, V.: Motor and cognitive functions of the ventral premotor cortex. Current Opinion in Neurobiology **12** (2002) 149–154
2. Touzet, C.: Neural reinforcement learning for behaviour synthesis. Robotics and Autonomous Systems **22** (1997) 251–81
3. Weber, C., Wermter, S., Zochios, A.: Robot docking with neural vision and reinforcement. Knowledge-Based Systems **17** (2004) 165–72
4. Jog, M., Kubota, Y., Connolly, C., Hillegaart, V., Graybiel, A.: Building neural representations of habits. Science **286** (1999) 1745–9
5. Elshaw, M., Weber, C., Zochios, A., Wermter, S.: A mirror neuron inspired hierarchical network for action selection. In: Proc. NeuroBotics. (2004) 89–97
6. Kohonen, T.: Self-Organizing Maps. Springer (2001)
7. Pulvermüller, F.: The Neuroscience of Language. On Brain Circuits of Words and Serial Order. Cambridge University Press (2003)
8. Kondo, T., Ito, K.: A reinforcement learning with evolutionary state recruitment strategy for autonomous mobile robots control. Robotics and Autonomous Systems **46** (2004) 111–24
9. Lee, I., Lau, Y.: Adaptive state space partitioning for reinforcement learning. Engineering Applications of Artificial Intelligence **17** (2004) 577–88
10. Barto, A., Mahadevan, S.: Recent advances in hierarchical reinforcement learning. Discrete Event Dynamic Systems **13** (2003) 41–77
11. Kalmár, Z., Szepesvári, C., Lörincz, A.: Module-based reinforcement learning: Experiments with a real robot. Machine Learning **31** (1998) 55–85
12. Humphrys, M.: W-learning: A simple RL-based society of mind. In: 3rd European Conference on Artificial Life. (1995) p.30
13. Knoblauch, A., Markert, H., Palm, G.: An associative model of cortical language and action processing. In: Proc. 9th Neural Comp. and Psych. Workshop. (2004)
14. Vitay, J., Rougier, N., Alexandre, F.: A distributed model of visual spatial attention. In: Biomimetic Neural Learning for Intelligent Robotics. Springer (2005)