Proc. of ICANN 2001.

Anticipation-Based Control Architecture for a Mobile Robot*

A. Heinze and H.-M. Gross

Ilmenau Technical University, Department of Neuroinformatics, P.O.B. 100565, D-98684 Ilmenau, Germany
{heinze,homi}@informatik.tu-ilmenau.de

Abstract. We present a biologically motivated computational model that is able to anticipate and evaluate multiple hypothetical sensorimotor sequences. Our Model for Anticipation based on Cortical Representations (MACOR) allows a completely parallel search at the neocortical level using assemblies of rate coded neurons for grouping, separation, and selection of sensorimotor sequences. For a vision-controlled local navigation of a mobile robot Khepera, we can demonstrate that our anticipative approach outperforms a reactive one. We also compare our explicitely planning approach with the implicitely planning Q-learning.

1 Introduction

Based on findings for the sensorimotor character of perception [1,2], we developed an alternative approach to perception that avoids the common separation of perception and generation of behavior and fuses both aspects into one consistent neural process. In this approach, perception of space and shape in the environment is regarded to be an active process which anticipates the sensory consequences of alternative hypothetical interactions with the environment, that could be performed by a sensorimotor system, starting from the current sensory situation. This approach is supported by biological findings. For example, it was shown that such planning of motor actions takes place in the secondary motor areas [3]. Thach (1996) found that the premotor parts of the brain are active both in planning movements to be executed as well as in thinking about movements that shall not be executed.

Based on these findings, we developed our Model for Anticipation based on Cortical Representations MACOR, presented in sec.2. It is intended as a general scheme for sensorimotor anticipation in a neural architecture. The model does not attempt to provide a detailed description of a specific cortical or subcortical structure, but we try to capture some general properties that are relevant to our "perception as anticipation"-approach in brain-like systems (for details see [4]). The objective of this paper is to demonstrate the efficiency of our anticipatory approach for a real-world sensorimotor control problem, the local navigation and obstacle avoidance of a vision-controlled mobile robot showing non-holonom movement characteristics. We compare the achieved navigation behavior with other non-anticipative approaches like reactive control and Q-learning-based control. It is important to note that the MACOR-concept is also suitable for other sensorimotor or cognitive tasks that must consider alternative action sequences and their hypothetical results.

* This work was partially supported by DFG Graduate college GRK 164.

2 Model for anticipation

In the framework of a visually-guided navigation task (sec.2.3), our architecture (see fig.1) processes as sensorimotor information a visual input yielded from an omnicamera, supplemented by the last motor command executed. Within the architecture, those inputs are represented by a Fuzzy ART architecture [5]. The visual part is represented by the F2-nodes of a Fuzzy ART network. Each of these nodes contains several neurons, which represent the motor commands available to bring the system into the considered visual situation (visuomotor column). Between the resulting visuomotor representation of the current situation and that of the preceeding one, connections incorporating the certainty of that transition and its evaluation are adapted (sec.2.1).

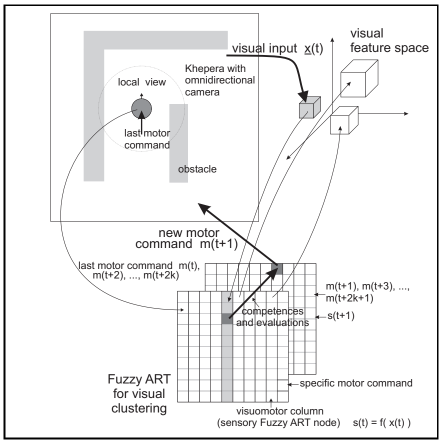

Fig. 1. Overview of the architecture. To represent visuomotor inputs, each Fuzzy ART node representing a specific visual input contains several neurons to represent motor commands available to bring the system in this situation. Connections between the present and the preceeding visuomotor situations are adapted to reflect the reinforcement incurred during that particular transition and the competence gained. By activation of a neuron and the subsequent spread of activity through its weights to activate other neurons, alternative sensorimotor sequences may be generated.

This architecture embeds the functionality required for a *parallel* generation of sequences of sensorimotor hypotheses (sec.2.2). In the context of the biological foundations of the architecture, this parallel generation is realized by a spread of activity through connections between the visuomotor columns. In the following investigations, we use a restricted set of discrete motor commands to investigate the effect of anticipative behavior (see sec.2.3, too).

2.1 Learning within the map

The learning of sensorimotor connections between the present and the preceding visuomotor situations takes place after each executed motor command. To evaluate the certainty of the existance of sensorimotor connections, we investigated several approaches, for example actual transition probabilities. Because of the clustering of the input space, high competence weights between neurons in the same sensorimotor assembly are established. The correspondingly lower competence weights between different situations yield lower sequence evaluations. These sequences are subsequentely not selected, but important for the consideration of movements resulting in a collision and thus for achieving of system goals.

Using the simple learning equ.1 realized as a sigmoidal function for **competence weights**, a specific weight can quickly reach a maximum value $w^{\mathrm{C}}_{\max}$, determined by the parameter μ and σ. Thus all weights hold the same value after a sufficiently high number of adaptation steps, such that the action selection (sec.2.2) depends only on the values of the evaluation weights. Therefore, the competence weights are only meaningful for unknown transitions, which are devalued by the competence. In equ.1, z_{ij} is the number of adaptations of the respective weight.

$$w_{ij}^{\mathrm{c}}(t) = \frac{1}{1 + e^{\frac{z_{ij} - \mu}{\sigma}}} \qquad \text{where} \qquad \mu = 2.8 \qquad \text{and} \qquad \sigma = 0.4 \tag{1}$$

For learning of **evaluation weights**, a simple form of reinforcement learning is used. Thus, the evaluation weights hold only the expected evaluation for the next transition, yielding a purely reactive behavior. Only by means of internal simulation, an anticipative behavior can be generated. The reinforcements r yield the evaluation of the generated behavior. In our experiments, they were chosen to reward a collision free, straight navigation behavior and to punish turns and collisions. For the adaptation of evaluation connections w_{ij}^{r}, the number z of adaptations of the respective weight w_{ij} is used to determine an adaptive learning rate $\frac{1}{z_{ij}^{r}}$.

$$w_{ij}^{\mathrm{r}}(\mathrm{t{+}1}) = w_{ij}^{\mathrm{r}}(\mathrm{t}) + \frac{1}{z_{ij}^{r}} \cdot (\mathrm{r} - w_{ij}^{\mathrm{r}}(\mathrm{t})) \tag{2}$$

For comparison, a special form of reinforcement learning, Q-learning [6, 7] was also investigated, because it is a model free, but implicitely planning approach.

$$w_{ij}^{\mathrm{r}}(\mathrm{t{+}1}) = w_{ij}^{\mathrm{r}}(\mathrm{t}) + \frac{1}{z_{ij}^{r}} \cdot (\mathrm{r} + \gamma \cdot w_{ki}^{\mathrm{r}}(t) - w_{ij}^{\mathrm{r}}(\mathrm{t})) \tag{3}$$

$$w_{ki}^{\mathrm{r}}(t) = \max_{l} w_{li}^{\mathrm{r}}(t) \tag{4}$$

By choosing the parameter γ to control the planning horizon $\gamma > 0$, the prediction of evaluations of subsequent transitions can be stored in one connection (strong Q-learning, equ.3 and 4). Q-learning becomes a planning free approach by choosing $\gamma = 0$ (weak Q-learning, equ.2).

2.2 Generation and evaluation of sensorimotor sequences

For generation of sensorimotor sequences, a specific neuron in a Fuzzy ART assembly is activated by the current sensory situation and the last executed motor command. This neuron propagates its activity y_j^{α} to all other interconnected neurons $i \in S = [0, n \cdot m - 1]$ using its competence connections w_{ij}^{c}, where n is the number of sensory assemblies and m the number of motor neurons within each assembly. The activated neurons may in turn activate further neurons, resulting in a mechanism of internal simulation and thus the generation of whole sequences of sensorimotor hypothesis (equ.5).

$$y_i^{\alpha}(t+1) = \max_{j \in S} w_{ij}^{\mathrm{c}}(t) \cdot y_j^{\alpha}(t) \tag{5}$$

Since the maximal value of the competence connection is less than 1.0, a subsequent neuron will always be less activated than its predecessor. Also, a neuron can only be activated by its maximum input activity. This supplies a simple stopping criterion for the propagation of interconnecting sequences. Simultaneous to the parallel generation of sequences of sensorimotor hypotheses, the model realizes the selection of the best evaluated sequence by a backpropagation of local sequence evaluations, as shown in equ.6. This backpropagation starts as the start neuron activates further neurons and each evaluation is backpropagated to the respective sequence predeccesor.

$$y_i^\beta(t+1) = w_{ij}^{\mathrm{r}}(t) \cdot y_i^\alpha(t+1) + \max_{r \in S} y_r^\beta(t) \tag{6}$$

The activity backpropagated to the start neuron represents the highest sequence evaluation in each time step. Thus in each time step, an action selection is possible, which improves as internal simulation goes on. This mechanism yields high cumulated sequence evaluations for well known and highly evaluated transitions. To realize a reactive behavior, the process of internal simulation runs only for one time step. This means the starting neuron activates further neurons which then propagate their evaluations directly back onto the starting neuron.

2.3 Experimental scenario

Because of their embodiment and situatedness, robots are ideal systems to demonstrate the advantages of an anticipation based sensorimotor control compared to a reactive one. To navigate successfully, for example, to avoid obstacles or to go through narrow passages, they have to consider their physical and mechanical properties and constraints (e.g. inertia, holonom or nonholonom kinematics). A mobile system, a robot or an animal, that is not able to learn and consider its constraints and their sensorimotor consequences, will not be able to evolve successful navigation behaviors.

To demonstrate the advantages of an anticipation based sensorimotor control, we used the mobile robot Khepera as a non-holonom system. Because of the used restricted action space, our system has to consider its constraints and is forced to start an early obstacle avoidance by internal simulation to realize a successful navigation behavior.

In our investigations, the visual sensory inputs were provided by an omnicamera. After a transformation into a physiological color space [8] specially tuned receptive fields extract the color distribution around the robot, which gives an implicit description of the obstacle arrangement.

3 Results

To demonstrate the advantages of anticipative systems, i.e. systems featuring explicit planning, we provide a comparison between the following approaches:

- the presented system which explicitely plans by internal simulation,
- a purely reactive system which operates exclusively on the current sensorimotor situation, and
- an implicitly planning approach, Q-learning.

For the investigation of anticipation-based behavior, we used Fuzzy-ART modules with 50 nodes (vigilance $\rho = 0.9$) for sensory clustering and three executable actions per sensory assembly. A sequential learning process was employed to structure the sensory clusters and to adapt the competence and evaluation weights involving more than 10.000 typical sensorimotor situations. In Q-learning we used a planning horizon of $\gamma = 0.9$. Systems except Q-learning, which does not use competence weights, utilized equ.1 to adapt their competence weights. The behavior generated by the different systems was recorded over 10 trials each and is exemplarily depicted in fig.2.

Fig. 2. In the experiments, a trial ended either after a collision or a maximum of 40 steps. The generated behavior shows that only the explicitly planning system (**middle**) was able to realize a successful obstacle avoidance considering the physical limitations of the non-holonom robot with a very limited action space. After the very first detection of a curve or obstacle, this system began to move to the left (as second selected action) and pursued that movement until the turn was negotiated. In contrast, the reactive system (**left**) chose to drive straight forward until almost hitting the wall. Only directly in front of the wall it chose to turn right, thus unable to avoid a collision. The Q-learning system (**right**) initiated also an early turn to the left and continued until the turn was negotiated. Subsequently, the action 'go straight' was repeatedly chosen, because the visual situation in the middle of the floor is represented by the same Fuzzy ART node as the floor situation near the wall. Only in the situation directly in front of the wall, another Fuzzy ART node was activated and a turn initiated. Like Q-learning, our reactive approach also learned the action 'go straight' for the floor situation, but by means of internal simulation, it chose mostly turns.

Further, in each trial the achievable total reinforcement was determined as the sum of the individual reinforcements received after each motor command. The respective mean values and variances together with typical trial lengths and the mean reinforcements for a single motor command are shown in table 1.

criteria		reactive system	explicitly anticip. system	Q-learning
reinforcement	mean	2.8	15.6	4.2
per trial	variance	0.1	8.6	2.4
trial length	mean	7.2	29.2	11.7
	variance	0.4	14.2	4.2
reinforcement per step	mean	0.61	0.55	0.57

Table 1. Mean reinforcement and standard deviation (std) for 10 trials of the different systems. Although the explicitly planning system achieves the highest mean reinforcement per trial, it also yielded the highest std. This was caused by 4 of those 10 trials (maximal trial length was 40) that ended due to collision, most probably as a matter of the hardware fixation of the omnicamera. The trials of the reactive and the Q-learning system all ended by collision and therefore in lower trial lengths. Further the mean reinforcements per step (not considering the step that resulted in a collision) show that the explicitely planning system got the lowest reinforcement per step as a result of more turns executed, which yield lower reinforcements than the movement 'go straight'.

These results give a strong indication that only the explicitly planning system is able to realize an obstacle avoidance for the inert, non-holonom roboter system. In comparison, the reactive system did not produce any avoiding actions at all. Though the Q-Learning system chose mostly proper motor commands, it could not overcome the perhaps non-optimal sensorimotor representation.

4 Conclusions and outlook

We presented MACOR that is able to anticipate and evaluate multiple hypothetical sensorimotor sequences. Using a framework of local navigation of a nonholonome robot, the advantage of anticipation based control was demonstrated empirically. The investigations performed demonstrate the advantage of anticipative systems compared to reactive systems and especially, the advantage of explicitly planning systems compared to implicitly planning ones.

This comparison is the basis of further investigations targeting a more complicated scenario, dealing with the aspect of changing system goals. These investigations are intended to show the advantages of reactively stored evaluation weights compared to weights whose values represent an implicit planning horizont, like Q-learning. These advantages become obvious for system goals that change over time, and which thus require readaptations of the sensorimotor transitions. With reactive evaluation weights in combination with internal simulation, only the actually different sensorimotor transitions must be relearned, which are only a few, to reflect the new system goal. So we want to show that an explicitly planning system may realize a flexible behavior according to a new task much faster than an implicitly planning system like Q-learning.

Further we want to investigate obvious similarities to probabilistic modelling techniques involving Hidden Markow Models [9] in detail. Besides some similarities in structure and purpose, there are a number of conceptual differences concerning, e.g., the learning strategies, the biological plausibility, or a prospective parallel feasibility.

References

1. M.A. Arbib, P. Erdi, and J. Szentagothai. *Neural Organization: Structure, Function and Dynamics.* MIT Press, 1998.
2. R. Pfeifer and C. Scheier. From Percept. to Action:The Right Direction? In *Proc.of PerAc'94*, pages 1–11. IEEE Comp.Soc.Press, 1994.
3. W. T. Thach. On the specific role of the cerebellum in motor learning and cognition: Clues from PET activation and lesion studies in man. *BBS*, 19(3):411–431, 1996.
4. H.-M. Gross, A. Heinze, T. Seiler, and V. Stephan. Generative Character of Percept.: A Neural Architecture for Sensorimotor Anticipation. *NN*, 12:1101–1129, 1999.
5. G. Carpenter et.al. Fuzzy ARTMAP: A Neural Netw. Architure for Incr. Supervised Learning of Analog Multidimensional Maps. In *NN*, pages 698–719, 1992.
6. C. Watkins. *Learn. from delayed rewards.* PhD thesis, King's Coll., Cambridge, UK, 1989.
7. R. S. Sutton and A. G. Barto. *Reinforcement Learning: An introduction.* MIT Press: A Bradford Book, Cambridge, MA, USA, 1998.
8. T. Pomierski and H.-M. Gross. Biological Neural Architecture for Chromatic Adaptation Resulting in Constant Color Sensations. In *Proc.IEEE Int. Conf. on NN*, pages 734–739, 1996.
9. L. R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. of IEEE*, 77(2):257–286, 1989.