A NEW APPROACH FOR SOLVING POSITIONING TASKS OF ROBOTIC SYSTEMS BASED ON REINFORCEMENT LEARNING

ALBERS, A[lbert]; SOMMER, H[ermann] & FRIETSCH, M[arkus]

Abstract: Automatic manipulators control poses a complex challenge for control systems, which have to deal with various dynamic and nonlinear effects. This paper presents a novel approach for motion control of a 2 DOF robot manipulator based on reinforcement learning. Through the implementation of the presented novel relative approach, it is possible to model every positioning task as a single offset compensation task. The experimental results for the completion of different positioning tasks using a single Q-Table are presented. The result is a considerably accelerated computation of the manipulators motion control.
Key words: Learning, Robotics, Dynamics, Control

1. INTRODUCTION

One of the biggest challenges in current research in robotics is, that robots "leave" their well structured environment and are confronted with new tasks in a more complex environment. An example for this new setting is the striving for autonomy and versatility in the fields of humanoid robotics as explained in (Peters J., Vijayakumar S. & Schaal S., 2003). Due to this, it can only be successful resp. useful, when it is able to adapt itself and to learn from its experiences. Reinforcement Learning (RL), a branch of machine learning (Mitchell, Tom M. 1997), is one possible approach to deal with this problem. However, the application of this learning process is limited due to its complexity. RL is a learning process, which uses reward and punishment signals from the interaction with the agent's environment in order to learn a distinct policy for achieving tasks. Various RL methods e.g. Q-learning (Watkins, 1989) or the SARSA algorithm have been studied in (Sutton, R.S. & Barto, A.G., 1998) where it is shown that two problems must be considered (Park, M. S. & Choi, J. Y. 2002). At first, the high computational efforts: RL is disturbed by the "curse of dimensionality" (Bellman, 1957), which refers to the tendency of a state space to grow exponentially in its dimension, that is, in the number of state variables.

Secondly, the information of every different task is stored in a separate Q-table, thus requiring a huge storing space when performing larger amounts of tasks. In addition, there is no formal connection between the Q-Tables corresponding to two different tasks; therefore the agent has to learn every task without any kind of precognition. This also reduces the usefulness of RL for practical applications and poses the question of how already acquired knowledge can be reused. In (Martin H, J. A. & De Lope, J., 2007), an approach is presented where a distributed architecture in RL serves as a pragmatical solution for the first problem for some common robotic manipulators with different DOF. A global high-dimensional Q-table which contains the evaluations of actions for all states is replaced by using some small, low dimensional Q-tables. RL based approaches have been applied to various robotic systems in the past, although mostly applied to the learning of elemental tasks meant to serve as "building blocks of movement generation" as in (Peters J., 2008). Nevertheless, new computations and additional storage space is required for performing new tasks. In this paper, a novel relative approach for positioning tasks is implemented, and the pragmatical repercussions resulting in a considerable reduction of the computational effort are presented.

1.2 Two Link Planar Robot

To show the feasibility of reinforcement learning methods to control manipulator systems, a 2 DOF manipulator system is used as a first simplified platform. Fig.1 shows the schematic drawing of the manipulator. The central goal of this paper is to move the robot manipulator from a start position to a target position. The aim is to create a trajectory using as few control commands as possible. Thus, the state of the manipulator is described by:

$$s = [\theta_1, \dot{\theta}_1, \theta_2, \dot{\theta}_2] \qquad (2)$$

The system is described more in detail in (Denzinger and Laureyns, 2008).
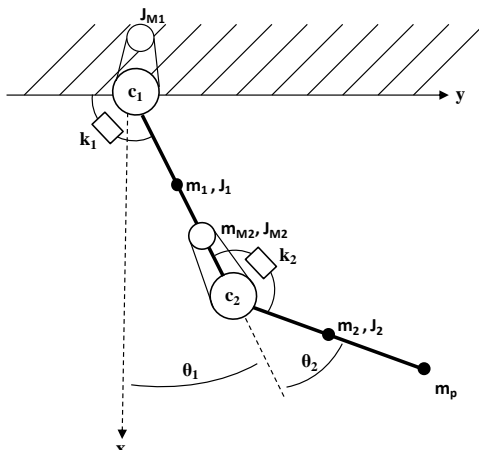

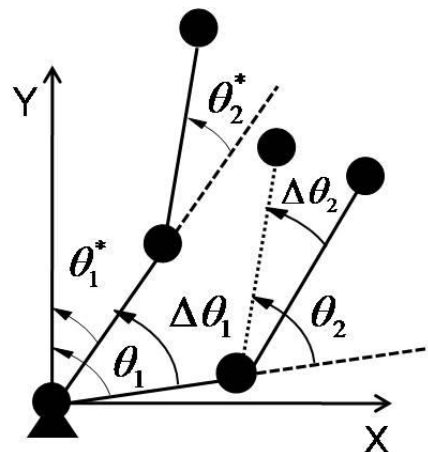Fig. 1. Dynamic modell of the manipulator

Fig. 1. Relative approach

2. THE RELATIVE APPROACH

2.1 Overview

The system is dynamic and nonlinear because the inertia of the upper arm changes due to a variable angle (Denzinger, J. & Laureyns, I. 2008). The state space of the robot arm consists of its angles θ and angle velocities $\dot{\theta}$. A Q-table in this case is a 6-dimensional hyperspace consisting of angles, velocities and actions:

$$Q = \{\theta_1, \theta_2, \dot{\theta}_1, \dot{\theta}_2, a_1, a_2\} \tag{1}$$

Applying the approach of (J. A. Martin H. & J. De Lope, 2007) delivers two low dimensional Q-table in $Q_1 = \{\theta_1, \dot{\theta}_1, a_1\}$ and $Q_2 = \{\theta_2, \dot{\theta}_2, a_2\}$. The relative approach developed in (Yan W. & et. al., 2009) enables each position to be calculated as a constant value plus a difference as shown in the follow equations:

$$(\theta_1, \theta_2) = (\theta_{1,target} + \Delta\theta_1, \theta_{2,target} + \Delta\theta_2)$$
$$(\Delta\theta_1, \Delta\theta_2) = (\theta_1 - \theta_{1,target}, \theta_1 - \theta_{1,target}) \tag{2}$$

At this point it is visible that every possible positioning task can be reduced to a simple offset compensation, reducing the amount of possible tasks to a single one, which contain all necessary information about the system's dynamics.

2.2 Computing appropriate Q-tables

As a result of the implemented relative approach, every positioning task can be seen as a compensation of an angular offset towards the target joint angle. Theoretically, the employed Q-Learning algorithm converges towards an optimal Q-table for the case that every state in the state space is visited an infinite amount of times. As a pragmatical solution to this problem the employment of an averaged Q-table for a finite amount of repetitions of the same task is proposed. Additionally, a proper reduction of the state space e.g. using an unevenly distributed state space makes it possible to visit all or nearly all states during a learning task.

3. RESULTS

The experimental focus lies in reusing the information of a learning task for the completion of new tasks. For this purpose a set of averaged Q-tables was computed. Each set corresponds to the average over a different amount of repetitions of one and the same learning task, consisting of 600 episodes with a maximum of 250 steps per episode. Here one completely performed movement starting from an initial position either ending after 250 steps or reaching the target position with less than 250 steps is termed as an episode. These Q-tables were computed for different amounts of repetitions using a Q-Learning algorithm. The computed Q-tables were then used to complete 10,000 random positioning tasks without further learning. The results are shown in Fig. 3. Using the averaged Q-tables for 10 and 100 repetitions ensured a perfect completion ratio. Furthermore, the quality of the solution increased with the amount of repetitions of the underlying learning task. The main difference between the employed policies could be detected at the amount of initialization values remaining in the Q-tables underlying each solution. While the Q-tables that corresponded to the solution computed for a single repetition of the learning task contained several initialization values, the ones corresponding to the average of one hundred repetitions presented only a single one left. These initialization entries re-

Repetitions	Completion ratio [%]	Average steps to target
1	99.2	65.55
10	100	31.28
100	100	26.99

Fig. 3. Experimental results

present situations in which the agent lacks information about its interaction with the environment, thus nearly always resulting in an inappropriate action selection and a failure of the exploitation task at hand. This pragmatical approach to reproduce the theoretical computation of an optimal policy failed to outperform a regular RL-algorithm in approximately one half of the tested tasks. The outcome of the comparison and the results are not surprising since an optimal policy cannot be computed as an average of non-optimal ones. The fact that the general quality of the solutions still improves with the amount of repetitions is a current topic of investigation.

4. CONCLUSIONS

An implementation of a novel relative approach for positioning tasks based on reinforcement learning was presented. By using this new approach any positioning task of a 2 DOF robot manipulator system can be modeled as an offset compensation task. This makes the employment of one single Q-table per joint, for the completion of any positioning task within the agent's workspace without additional learning possible. At this point no further computation is necessary to find a solution for any new positioning task, which represents a vast reduction of the computing costs for the control of the manipulator. The validation of the results obtained by the conducted simulations using a real manipulator system will be subject to future work. In general this approach can be seen as a self learning regulator framework for closed loop control of diverse nonlinear systems, which will also be subject to future research.

5. REFERENCES

Bellman R. (1954). The theory of dynamic programming, *Bulletin of the American Mathematical Society*, 60, 503-515

Denzinger J.; Laureyns I. & et.al. (2008). A study of reward functions in reinforcement learning on a dynamic model of a two-link planar robot, The 2nd European DAAAM International Young Researchers´ and Scientists´ Conference

Martin J. A. & De Lope J. (2007). A Distributed Reinforcement Learning Architecture for Multi-Link Robots, *4th International Conference on Informatics in Control, Automation and Robotics* (ICINCO), (pp. 192-197). Angers, France

Mitchell, Tom M. (1997). *Machine learning*, McGraw-Hill, New York, ISBN 0-07-042807-7

Peters J., Vijayakuar S. & Schaal S. (2003) Reinforcement learning for humanoid robotics. Third IEEE-RAS International Conference on Humanoid Robots, Karlsruhe, Germany.

Peters J. (2008). Machine Learning for Robotics. VDM Verlag Dr. Müller, Saarbrücken, ISBN 978-3-639-02110-3

Sutton R.S & Barto A.G. (1998). Reinforcement Learning, an introduction, *The MIT press*, MA

Yan, W & et. al. (2009). Application of reinforcement learning to a two DOF Robot arm control, Annals of DAAAM for 2009 & Proceedings of 20th DAAAM International Symposium