A Qualitative Representation of Structural Spatial Knowledge for Robot Navigation with Reinforcement Learning

Lutz Frommberger LUTZ@SFBTR8.UNI-BREMEN.DE
SFB/TR 8 Spatial Cognition, R3-[Q-Shape], Universität Bremen, P.O. Box 330 440, 28334 Bremen, Germany

Abstract

In robot navigation tasks, the representation of knowledge of the surrounding world plays an important role, especially in reinforcement learning approaches. This work presents a qualitative representation of space that empowers an agent to learn a goal-directed navigation strategy based on structural knowledge of the world that leads to a generally sensible navigation behavior that can be transferred to completely unknown environments.

1. Introduction

In goal-directed navigation tasks, an autonomous moving agent has fulfilled its mission when having reached a certain location in space. Reinforcement Learning (RL) is frequently applied to such tasks, because it allows an agent to autonomously adapt its behavior to a given environment. However, in large and in continuous state spaces RL methods require extremely long training times.

The navigating agent learns a strategy that will bring it to the goal from every position within the world, that is: it learns to select an action for every given observation of the environment. But what has the agent really learned about the world it operates in? Can it use the acquired knowledge at different locations in this environment or even in an unknown one? This depends heavily on the chosen spatial representation of the world that is passed to the learning system. In the worst case, all the collected knowledge can become useless. For example, when the agent operates in a different environment, a completely new set of action selections may be necessary, and the agent has to learn everything again from scratch, including collision avoidance strategies. The agent lacks an *understanding* of the general structure of geometrical spaces.

Thrun and Schwartz (1995) claim that it is necessary to discover the structure of the world and abstract from its details to be able to adapt RL to more complex tasks. The aim of the approach I present in this paper is to enable the agent to profit from this structure by using an appropriate *qualitative* representation for it. Qualitative spatial representations provide an interface that is based on human spatial concepts. They are an expressive means of representing the relations among features in geometrical space. I claim that this way of representing structural spatial information can enable the agent to develop a generally sensible behavior in space that it can reuse at different locations within the same world and can also be transferred to learning tasks in other, unknown environments. The use of this representation will also support the learning process by speeding it up and making it more robust.

Much effort has been spent to accomplish improvements regarding the training speed of reinforcement learning in navigation tasks, and consideration of the structure of the state space has been found to be an important means to reach that goal. Topological neighborhood relations can be used to improve the learning performance (Braga & Araújo, 2003). Thrun and Schwartz (1995) tried to find reusable structural information that is valid in multiple tasks. Glaubius et al. (2005) concentrate on the internal value-function representation to reuse experience across similar parts of the state space, using pre-defined equivalence classes. Lane and Wilson (2005) describe relational policies for spatial environments and demonstrate significant learning improvements. To avoid problems with walls, they also suggest regarding the relative position of walls with respect to the agent, but did not realize this approach yet.

The first goal of this work is to find a spatial representation that leads to a small and discrete state space which enables fast and robust learning of a navigation strategy in a continuous, non-homogeneous world.

Appearing in the ICML-06 Workshop on Structural Knowledge Transfer for Machine Learning, June 29, Pittsburgh, PA.

The second goal is to extract structural knowledge from the environment to enable the agent to reuse learned strategies in similar areas within the same world and also being able to transfer this knowledge to other, unknown environments.

2. Navigation solely on the basis of ordering of landmarks

The task considered within this work is a simple goal-directed navigation task: an autonomous robot is requested to find a certain wall in a simulated simplified office environment completely unknown to the agent (see Figure 1). The robot is supposed to be capable to determine landmarks around it to identify its location. It is assumed that every wall is uniquely distinguishable, making the whole wall a landmark of its own. To represent this, each wall is considered to have a certain color.

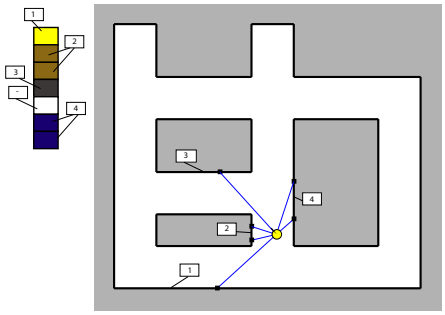

Figure 1. The task: a robot in a simplified simulated office environment with uniquely distinguishable walls. Detected colors are depicted in the upper left, with labels attached marking the corresponding walls.

The robot is capable of performing three different basic actions: moving forward and turning a few degrees both to the left and to the right. Each basic action is repeated as long as the perception vector remains the same. There is no built-in collision avoidance or any other navigational intelligence provided. A small amount of noise is added to the motor actions.

The agent uses a compact *qualitative abstraction* of real world information: the colors perceived at 7 discrete angles around it, a vector $c = (c_1, \ldots, c_n)$. Every physical state, a tuple (x, θ) of the robot's position x and orientation θ, maps to exactly one color vector c. This mapping is not injective: multiple physical states share the same representation. The encoding of a circular order of perceived colors is sufficient to approximately represent the position of the agent within the world and to derive a sequence of actions to reach the goal. However, it is not sufficient to prevent the robot from collisions. As stated above, the mapping from physical locations to the state representation is not unique, and, given the same system input, the consequences of an action can differ dramatically. Performing the same action at the same system state will sometimes result in a collision and sometimes not, which prevents from retrieving a proper rating for this state-action pair. The representation used so far does not encode any information about the agent's position regarding the obstacles. It does not support the agent in developing a generally sensible spatial behavior that should also emerge in the absence of landmarks.

3. (G)RLPR: a spatial representation of relative position of line segments

Navigation in space, as performed in the learning examples, can be viewed as consisting of two different aspects: (1) *Goal-directed behavior* towards a certain target location depends highly on the task that has to be solved. If the task is to go to a certain location, the resulting actions at a specific place are generally different for different targets. Goal-directed behavior is task-specific. (2) *Generally sensible behavior* regarding the structure of the environment is more or less the same in identical or similar environments. A generally sensible behavior in indoor office environments for example would be not to crash into walls, turn around corners smoothly etc. It does not depend on a goal to reach, but on characteristics of the environment that invoke some kind of behavior. Generally sensible behavior is task-independent. The aim is to find a representation that divides between the two aspects of navigation behavior in order to be able to extract generally sensible behavior from a learned strategy.

The relations of walls towards each other induce sensible paths inside the world which the agent should learn to follow. I propose a representation of the relative positions of lines towards the agent's moving direction. For further reference, it is called *RLPR* (Relative Line Position Representation). RLPR is intentionally chosen to be extremely compact while being very expressive to be added to an existing feature vector.

To encode the relative positions of certain entities regarding the agent's position and orientation, we construct an enclosing box around the robot and then extend the boundaries of this box to create eight disjoint regions R_1 to R_8 (see Figure 2a). This representation was proposed to model the movement of extended objects in a qualitative manner (Mukerjee & Joe, 1990) and is closely related to the *direction-relation matrix* (Goyal & Egenhofer, 2000). We can define a *traversal status* $t(B, R_i)$ of each line B regarding a region R_i as

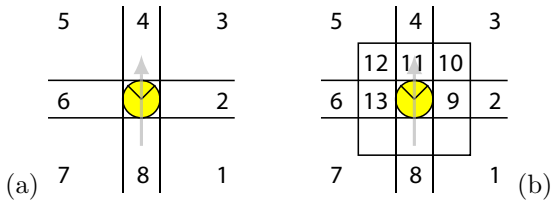

Figure 2. Neighboring regions around the robot in relation to its moving direction. The regions in the immediate surroundings (b) are proper subsets of $R_1, \ldots, R_8$ (a).

follows: $t(B, R_i) = 1$ if a line B cuts region R_i and 0 if not. The total number of lines in a region R_i is

$$t(R_i) = \sum_{B \in \mathcal{B}} t(B, R_i) \tag{1}$$

with $\mathcal{B}$ being the set of all detected line segments.

For navigation purposes, it is particularly interesting to know which line segments span from one region to another. A corridor with a left turn, for example, will have connected line segments in all the right and front sectors, but none in freely traversable space. To additionally encode that a line segment B spans from one sector to another, we determine if a line B lies within counter-clockwise adjacent regions R_i and R_{i+1} (for R_8, of course, we need to consider R_1):

$$t'(B, R_i) = t(B, R_i) \cdot t(B, R_{i+1}) \tag{2}$$

$t'(B, R_i)$ is also very robust to noisy line detection, as it does not matter if a line is detected as one or more segments. The total number of spanning line segments in a region, $t'(R_i)$, is derived analogously to (1).

To achieve a valuable representation of the environment, special care has to be taken on the immediate surroundings of the agent. The position of detected line segments is interesting information to be used for general orientation and mid-term planning. Additionally, if an object is detected in the immediate surroundings, this information refers to obstacle avoidance and requires a prompt reaction. It can also be utilized to realize certain behaviors, for example, a wall following strategy. For regarding the regions near the agent, the representation described above is used twice. On the one hand, there are the regions $R_1, \ldots, R_8$ that are bounded by the perceptual capabilities of the robot. One the other hand, bounded subsets of those regions represent the immediate surroundings (see Figure 2b).

To combine knowledge about goal-directed and generally sensible spatial behavior, we now build a feature vector by concatenating the representation of detected colors and RLPR. A system state s is represented as

$$s = (c_0 \ldots c_n, t'(R_1) \ldots t'(R_8), t(R_9) \ldots t(R_{13})) \tag{3}$$

Within the immediate surroundings, the information of spanning line segments is not that important, so $t(R_i)$ is used instead of $t'(R_i)$ for $i \geq 9$.

In the following I want to show that while learning a goal-directed navigation task with a clearly specified target location the agent can also learn a sensible behavior regarding the structure of the surrounding world independent of landmark information when using RLPR. Therefore, the generalization abilities of *tile coding* (Sutton, 1996) are used in the value function representation over the complete range of color information. This means that an update of the policy affects all system states with the same RLPR representation. This generalizing variant of RLPR is called *Generalizing RLPR* (GRLPR). As an effect, the agent can reuse knowledge acquired within the same learning task. The learned strategy is also applicable to new environments that are completely unknown to the agent.

4. Experimental results

All experiments have been conducted using Watkins' Q(λ) algorithm. In a first experiment, the robot has to solve the goal finding task in the world depicted in Figure 1. When using only $(c_1, \ldots, c_7)$ as the input ("pure" approach), no robust learning behavior can be achieved. In contrast, the additional use of RLPR or GRLPR leads to stable learning: after about 10,000 episodes, almost all of the test runs reach the goal.

Figure 3 compares the "pure" approach with RLPR and GRLPR during the first 15,000 episodes. Due to the non-generalizing behavior and the larger feature vector compared to the "pure" approach, RLPR shows a slower learning in the very beginning, but gets comparably successful as GRLPR over time. GRLPR learns faster than the other two approaches in the early training phase. It benefits from its generalizing behavior that empowers it to reuse structural spatial knowledge gained in other parts of the environment.

An important measure of the quality of navigation is the number of collisions that occur during training. This number is reduced significantly with the additional line position representations (see Table 1). In particular, GRLPR performs noticeably better than RLPR. This indicates that the generalization ability leads to a sensible navigation behavior rather early.

To show that the learned knowledge of the structure of the world is not only beneficial within the same learning task, we must examine how the agent behaves when using the learned strategy in the absence of landmarks or in an unknown world. After learning with GRLPR for 40,000 episodes, the landmark information

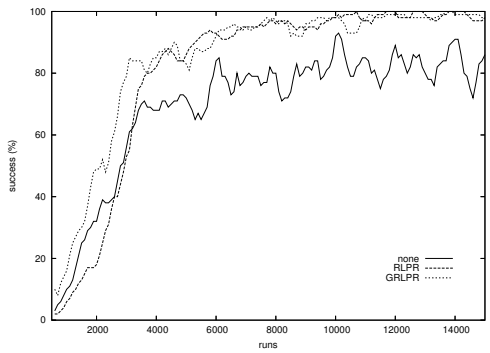

Figure 3. Number of runs reaching the goal state: GRLPR learns faster than the "pure" approach. RLPR is slower in the beginning, but catches up quite fast. Without (G)RLPR, no stable learning can be achieved.

Table 1. Number of collisions after 15,000 episodes

Repr.	Training	Test
pure	21388	2385
RLPR	11316	1200
GRLPR	7929	596

is turned off, so that the agent perceives the same (unknown) color vector regardless of where it is. As a result, the agent is still able to navigate collision-free and perform smooth curves, even without receiving any landmark information. Most of the time it uses a follow-the-wall strategy, a commonly-used strategy in robotics. The learned strategy can also successfully be transferred to absolutely unknown environments. Figure 4 shows the agent's trajectories in a landmark-free world it has never seen before, using the gained knowledge from the prior experiment.

5. Conclusion and outlook

Solving a goal-directed robot navigation task can be learned with reinforcement learning using a qualitative spatial representation purely using the ordering of landmarks and the relative position of line segments. The proposed representation results in a fast and stable learning of the given task. Structural information within the environment is generalized and can be reused within the same learning task, which leads to a generally sensible navigation strategy that facilitates a faster learning and reduces collisions significantly. Furthermore, the learned knowledge can be transferred to environments lacking landmark information and/or totally unknown environments: the agent learns a generally sensible behavior in geometrical spaces.

In future work it has to be shown that the acquired

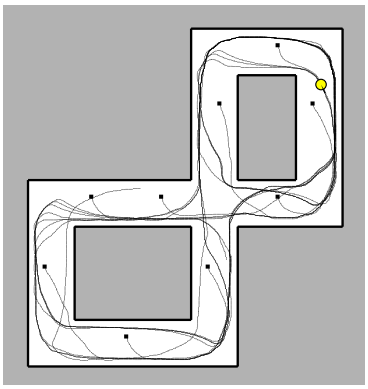

Figure 4. Trajectories of the agent in an unknown environment without landmarks, using the strategy learned in a different world with GRLPR.

strategy, used as background knowledge, can speed up new learning tasks in unknown environments. Also other methods of evaluating structural knowledge, e.g., in a hierarchical manner, need to be investigated.

References

Braga, A. P. S., & Araújo, A. F. R. (2003). A topological reinforcement learning agent for navigation. *Neural Computing and Applications*, *12*, 220–236.

Glaubius, R., Namihira, M., & Smart, W. D. (2005). Speeding up reinforcement learning using manifold representations: Preliminary results. *Proceedings of the IJCAI Workshop "Reasoning with Uncertainty in Robotics".*

Goyal, R. K., & Egenhofer, M. J. (2000). Consistent queries over cardinal directions across different levels of detail. *Proceedings of the 11th Intl. Workshop on Database and Expert System Applications.*

Lane, T., & Wilson, A. (2005). Toward a topological theory of relational reinforcement learning for navigation tasks. *Proceedings of FLAIRS-2005.*

Mukerjee, A., & Joe, G. (1990). A qualitative model for space. *Proceedings of the Eighth National Conference on Artificial Intelligence (AAAI).*

Sutton, R. S. (1996). Generalization in reinforcement learning: Successful examples using sparse tile coding. In D. S. Touretzky, M. C. Mozer and M. E. Hasselmo (Eds.), *Advances in neural information processing systems*, vol. 8, 1038–1044.

Thrun, S., & Schwartz, A. (1995). Finding structure in reinforcement learning. In G. Tesauro, D. Touretzky and T. Leen (Eds.), *Advances in neural information processing systems*, vol. 7.