Interaktive Objekterkennung – Combining Reinforcement Learning and Belief Revision – A Learning System for Active Vision

ProjektleiterInnen
Prof. Dr. Gabriele Peters
Prof. Dr.
Gabriele Kern-Isberner,
Technische Universität
Dortmund,
Fakultät für Informatik

Zeitraum
2005 – 2007

Kontakt
Prof. Dr. Gabriele Peters
Fachbereich Informatik
Fachhochschule
Dortmund
Emil-Figge-Straße 42
44227 Dortmund
Tel.: (0231) 755-6796
E-Mail: gabriele.peters
@fh-dortmund.de

Abstract

Computer vision can highly benefit from modern learning methods. In the context of an active vision environment we develop a machine learning approach which is able to learn strategies of object acquisition. We propose a hybrid learning method, called Sphinx, that combines two approaches originating from seperate disciplines of computer science, namely reinforcement learning on the one hand and belief revision on the other. The former represents knowledge in a numerical way, while the latter is based on symbolic logic and allows reasoning. Sphinx is designed according to human cognition and interacts with its environment by rotating objects depending on past perceptions to acquire those views which are advantageous for recognition. Our method was successfully applied in simulations of object categorization tasks.

1 Introduction

One of the most challenging tasks of computer vision systems is the recognition of known and unknown objects. An elegant way to achieve this is to show the system some samples of each object class and thereby train the system, so that it can recognize objects that it has not seen before, but which look similar to some objects of the training phase (due to some defined features). For this purpose several methods have been successfully used and analyzed. One of them is to set up a rule-based system and have it reason, another one is to use numerical learning methods such as reinforcement learning. Both of them have advantages, but also disadvantages. Reinforcement learning yields good results in different kinds of environments, but its training is time consuming, since it is a trial-and-error method and the agent has to learn from scratch. The possibilities to introduce background knowledge (e.g., by the choice of the initial values of the Q-table) are more limited as for example with knowledge representation techniques. Another disadvantage consists in a limited possibility to generalize experiences and thus to be able to act appropriately in unfamiliar situations. Though some generalization can be obtained by the application of function approximization techniques, the possibilities to generalize from learned rules to unfamiliar situations are more diverse again with knowledge representation techniques. Knowledge representation and belief revision techniques have the advantage that the belief of the agent is represented quite clearly and allows reasoning about actions. The belief can be extended by new information, but needs to be revised when the new information contradicts the current belief. One drawback is that it is difficult to decide which parts of the belief should be given up, so that the new belief state is consistent, i.e., without inherent contradictions. In this project we analyse the hybrid learning model Sphinx, named after the Egyptian statue of a hybrid between a human and a lion. It combines the advantages of both approaches and diminishes the disadvantages, thus synergy effect can emerge. Summarizing, the problem we address is not object recognition, rather we propose a machine learning system which interacts with its environment and is able to learn autonomously strategies to explore the environment, i.e., to learn object acquisition strategies. To demonstrate the capabilities of Sphinx we test it – in a coarse manner – in a first object classification task.

2 Related Work

Psychological findings propose a two-level learning model for human learning. On the so called bottom level, humans learn implicitly and acquire procedural knowledge. They are not aware of the relations they have learned and can hardly put it into words. On the other level, the top level, humans learn explicitly and acquire decalarative knowledge. They are aware of the relations they have learned and can express it, e.g., in form of if-then rules. A special form of declarative knowledge is episodic} knowledge. This kind of knowledge is not of general nature, but refers to specific events, situations or objects. Episodic knowledge facilitates to remember specific situations where general rules do not apply. These two levels do not work separately. Depending on what is learned, humans learn top-down or bottom-up. It has been found that in completely unfamiliar situations mainly implicit learning occurs and procedural knowledge is acquired. The declarative knowledge is formed afterwards. This indicates that the bottom-up direction plays an important role. It is also advantageous to continually verbalize what has been learned and thus speed up the acquisition of declarative knowledge. Sun, Merrill and Peterson developed the learning model CLARION. It is a two-level, bottom-up learning model which uses Q-learning for the bottom level and a set of rules for the top level. The rules have the form ‚Premise => Action‘, where the premise can be met by the current state signal of the environment. For the maintainance of the set of rules (i.e., adding, changing and deleting rules) the authors have developed a certain technique. They have proven their model, which works similar to human learning, to be successful in a mine field navigation task.

3 The Sphinx Learning Approach

Like the CLARION system, our learning approach consists of two levels. For the bottom level we use Q-Learning. For the top level we utilize belief revision techniques and an ordinal conditional function (OCF) to represent the epistemic state of an agent. Ordinal conditional functions are also called ranking functions, as they assign a degree of disbelief or surprise to each model. In this way we obtain a framework that has a well analysed, theoretical background.
In the following we will briefly decribe Q-learning. The setting consists of an environment and one or more agents. An agent can interact with the environment. Usually, the environment starts in a state and ends, when one terminal state is reached. This timespan is called an episode. For each action, the agent is rewarded and it aims at collecting high rewards during an episode. Episodes consist of steps. A step is the following: The agent perceives the current state of the environment via a (numerical) state signal, e.g., an ID. It looks up that action in its memory (for Q-Learning this is expressed be the Q-Table), which seems to be the best in this situation and performs it. The environment reacts on this action by changing its state. After this change, the agent is rewarded for its action and updates its Q-table.
To combine belief revision and reinforcement learning, we added a symbolic (i.e., a conjunction of literals) representation of the states of the environment, thus states s have a dual representation: a numerical one for the reinforcement learning part and a symbolic one for the belief revision part. The Sphinx system is displayed in figure 1 and works as follows:

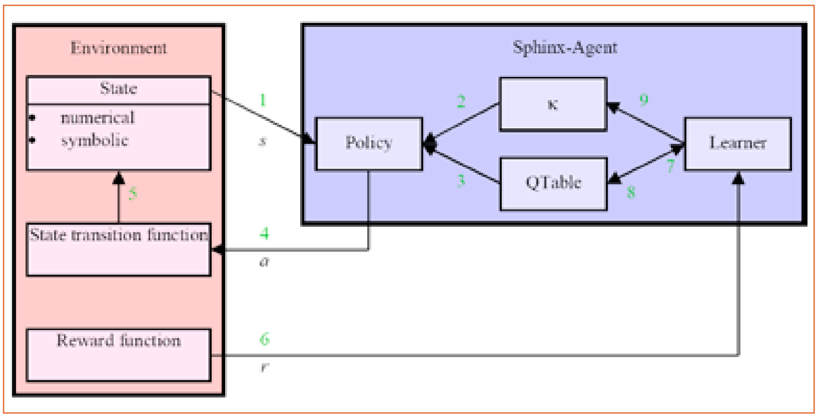

Figure 1. Sphinx Learning Approach.

Algorithm ‚Sphinx-Learning‘:

1. The Sphinx agent perceives the signal of the dually represented state s coming from the environment.
2. The agent queries its OCF about which actions are most plausible in s.
3. The agent looks up the Q-values of these actions and determines the set of those actions that have the greatest Q-value. (An ordinary Q-agent determines the set of best actions from the set of all possible actions.)
4. The agent chooses a random action and performs it.
5. The environment changes to the successor state.
6. The agent receives the reward r.
7. The agent updates the Q-table.
8. The new Q-values for actions in s are being read.
9. The agent revises the OCF. (This revision usually makes those actions most plausible in s that have the greatest Q-value in s. In addition, the agent tries to find patterns in the state signals for which certain actions are generally better than others.

Since revisions and especially revisions with generalized rules have a strong influence on the choice of actions, they have to be handled carefully, i.e., the agent should be quite sure about the correctness of a rule before adding it to its belief. Therefore, the agent uses several counters counting, how ‚often‘ an action has been poor, not poor, a best or not a best one under certain circumstances. With these counters probabilities can be calculated which can be used to evaluate the certainty about the correctness of a specific rule.
Our learning model also supports background knowledge. If the user knows some rules that might be helpful for the agent and its task, she can formulate them as conditionals and let the agent revise the OCF with them before starting to learn.

4 Recognition of Simulated Real Objects

We defined shape attributes that are suitable for representing objects in a simple object recognition task and then chose arbitrary objects and described them with these previously defined attributes. These attributes are the input to Sphinx. There are three possible perspectives: the front view, the side view, and a view from a position between these two views. The front and the side view are described by three attributes each: approximate (idealized) shape, size (i.e., proportion) of the shape, and deviation from the idealized shape. The approximate shape can take the values unknown, circle, square, triangle up, and triangle down. The size can be unknown, flat, regular, or tall. The deviation can be small, medium, or large. Besides these attributes the object is described by the complexity of its texture. This attribute can take the values simple, medium, and complex. We set the attributes for each object manually. If the agent looks at the object from the front or the side, it perceives the matching idealized shape, its size, its deviation, and the complexity of the texture. From the intermediate view the agent can only perceive the idealized shapes of the front and the side view and the complexity of the texture, but not the size and deviations. Possible actions of the agent are rotate left, rotate right, recognize unknown and recognize x, where x is one of the objects. At the beginning of each episode, the agent looks at the current object from a random perspective and the variables are set according to this perspective. Now the agent can rotate the object clockwise or counter-clockwise or name it. If the agent's action is a ‚recognize...' action, the episode ends. After ten steps the running episode is forced to end. We have chosen 15 different objects from nine different object classes such as bottle, tree, and house for which we provide the three attributes mentioned (shape, size, and deviation) (figure 2). Figure 3 shows the results averaged from 100 independed agents.

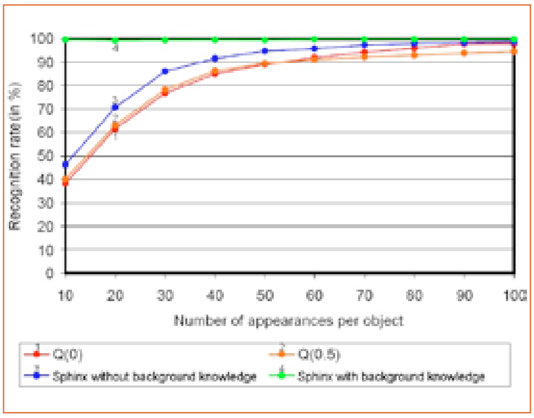

Figure 3. Recognition rates for simulated real objects.

In a second step we added background knowledge that enables the agent to recognize all objects correctly, if it has perceived all of the three views. Furthermore, we added rules to the background knowledge that told the agent to look at the object from all perspectives first. With these rules the agent has a complete, but not optimal, solution for the task. We wanted to find out how fast the agent learns that it does not need all views to classify the current object. This setup resulted in a constantly high recognition rate from 99.3% to 99.8%. The number of perceived views decreased over time from 3.28 to 1.99 (figure 3).

Figure 2. Two examples from the object classes bottle and house. The two left hand images show the front and side view of a bottle with medium texture. The two right hand images show the front and side view of a house. Shape, size, and deviation of the front view are square, regular, and little, of the side view triangle up, regular, and medium, respectively.

Here are some of the rules the agent learned and assimilated during its training:

If FrontViewShape = triangle up and FrontViewSize = tall, then action = recognize bottle.

If FrontViewShape = circle and SideViewShape = unknown and texture = simple, then action = rotate left.

If Texture = complex, then not (action = recognize bottle).

5 Conclusion

Summarizing one can say that Sphinx has an advantage over classic reinforcement learning in terms of learning speed by combining two cognitive levels of learning – similar to human learning. This advantage remains even if Sphinx is not provided with background knowledge., i.e., when both learning methods are started equivalently. Nevertheless, background knowledge can support the process of learning, so that even at the beginning high success rates can be achieved. It can be described in a manner comprehensible for humans in the form of conditionals. This is not feasible in classical reinforcement learning. Moreover, learned information can be output easily comprehensible for humans in the form of rules; there is no need to interprete numerical data. By providing Sphinx with background knowledge in the form of obvious (but not necessarily optimal) rules permanently high learning rates can be obtained, while the amount of required information to choose a proper action decreases. These characteristics qualify our hybrid learning method especially to be applied in active vision environments.

References

- Thomas Leopold, Gabriele Kern-Isberner, and Gabriele Peters, Belief Revision with Reinforcement Learning for Interactive Object Recognition, 18th European Conference on Artificial Intelligence, (ECAI 2008), 2008.
- Thomas Leopold, Gabriele Kern-Isberner, and Gabriele Peters, Combining Reinforcement Learning and Belief Revision – A Learning System for Active Vision, submitted to: 19th British Machine Vision Conference, (BMVC 2008}, 2008.