ARTICLE IN PRESS

ARTINT:2827

Artificial Intelligence ••• (••••) •••–•••

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Relational reinforcement learning with guided demonstrations

David Martínez*, Guillem Alenyà, Carme Torras

Institut de Robotica i Informatica Industrial (CSIC-UPC), Llorens i Artigas 4-6, 08028 Barcelona, Spain

ARTICLE INFO

Article history:
Received in revised form 15 February 2015
Accepted 16 February 2015
Available online xxxx

Keywords:
Active learning
Learning guidance
Planning excuse
Reinforcement learning
Robot learning
Teacher demonstration
Teacher guidance

ABSTRACT

Model-based reinforcement learning is a powerful paradigm for learning tasks in robotics. However, in-depth exploration is usually required and the actions have to be known in advance. Thus, we propose a novel algorithm that integrates the option of requesting teacher demonstrations to learn new domains with fewer action executions and no previous knowledge. Demonstrations allow new actions to be learned and they greatly reduce the amount of exploration required, but they are only requested when they are expected to yield a significant improvement because the teacher's time is considered to be more valuable than the robot's time. Moreover, selecting the appropriate action to demonstrate is not an easy task, and thus some guidance is provided to the teacher. The rule-based model is analyzed to determine the parts of the state that may be incomplete, and to provide the teacher with a set of possible problems for which a demonstration is needed. Rule analysis is also used to find better alternative models and to complete subgoals before requesting help, thereby minimizing the number of requested demonstrations. These improvements were demonstrated in a set of experiments, which included domains from the international planning competition and a robotic task. Adding teacher demonstrations and rule analysis reduced the amount of exploration required by up to 60% in some domains, and improved the success ratio by 35% in other domains.

1. Introduction

Learning tasks with robots is a very interesting topic, where impressive results may be obtained without having to design specific algorithms for each task. However, learning high-level tasks can be very time consuming because considerable experience is required to learn in real-world domains where stochastic actions with multiple effects can be performed. In general, learning from scratch in such domains requires that hundreds of actions are executed even for simple tasks. Robots require several seconds, or even minutes, to execute high-level actions, which means that the total time required to learn a task can be excessively large. However, learning with the help of a human teacher can reduce this learning time greatly, although human time is usually considered to be more valuable than that of a robot, and thus the robot should only request help from a teacher a limited number of times.

This is the underlying concept of the learning approach in the European project IntellAct [1], the goal of which is to exploit the semantics of manipulations in terms of objects, actions, and their consequences to reproduce human actions using robots. This project provides a framework where a robot should learn manipulation tasks with no prior knowledge, as well as performing high-level reasoning to adapt to changes in the domain. The robot starts without any knowledge of

* Corresponding author.
E-mail addresses: dmartinez@iri.upc.edu (D. Martínez), galenya@iri.upc.edu (G. Alenyà), torras@iri.upc.edu (C. Torras).

http://dx.doi.org/10.1016/j.artint.2015.02.006
0004-3702/

Please cite this article in press as: D. Martínez et al., Relational reinforcement learning with guided demonstrations, Artificial Intelligence (2015), http://dx.doi.org/10.1016/j.artint.2015.02.006

the available actions and it has to request demonstrations from a teacher whenever new actions are required to complete a task. These actions are learned at a low level using programming by teleoperation [2], as well as at a high level with a decision maker. Moreover, a 3D object recognition algorithm [3] and a particle filter tracker [4] are used to obtain a state for reasoning and to generate semantic event chains [5], which encode the demonstrated action sequences. Using this information, the decision maker can learn about the domain constantly while completing the assigned tasks [6]. In this study, we propose a new reinforcement learning algorithm for the decision maker called Relational Exploration with Demonstrations (REX-D), which can learn to perform manipulation tasks based on only a small number of action executions with the help of a teacher guided by rule analysis.

In some robotic systems, such as those that involve complex manipulation sequences, the decision maker has little input data and long periods of time to process it, because the actions take a long time to execute. Therefore, a good approach for learning such high-level tasks is model-based reinforcement learning (RL) [7]. This approach allows a model to be obtained that represents the actions that the robot can execute. The model is generated from the experiences obtained when the robot executes the actions.

To learn tasks as rapidly as possible, we need a highly compact representation of the model, and thus we use relational models. These models generalize over different objects of the same type, thereby reducing the learning complexity of domains to the number of different types of objects in them. Several approaches apply RL to obtain good results in specific robotic tasks using relational domains [8,9], as well as with general relational models [10,11]. A fundamental problem in RL is balancing exploration and exploitation. To reduce the number of experiences required, Lang et al. [12] proposed the REX algorithm, which used relational count functions to apply relational generalization to the exploration–exploitation dilemma. In our approach, we learn general relational models using the relational generalization of REX, but we also include teacher demonstrations and rule analysis to reduce the number of experiences required even further, as well as facilitating generalization over different tasks.

In general, there is some uncertainty about the effects of an action executed by a robot. This uncertainty is important in some tasks, thus the RL algorithm should be able to handle stochastic effects. In the KWIK framework [13], a method was proposed for learning the probabilities associated with a given set of action effects using linear regression [14], as well as an extension for learning the action effects themselves [15]. However, a large number of samples are needed because the problem of learning action effects is NP. In our proposed method, we use the learner proposed by Pasula et al. [16], which employs a greedy algorithm to obtain rule sets that optimize a score function. Although this does not obtain the optimal solution, it generates good rule sets based on only a few experiences. Furthermore, it generates rules with deictic references and noisy effects, which make models more compact and tractable.

Learning from Demonstration (LfD) is a supervised learning paradigm, where a teacher transfers knowledge of tasks or skills to a robotic system by performing a demonstration of the task or skill [17]. Thus, we propose the combination of RL with LfD to obtain a system that learns new tasks without needing to know the actions in advance. Demonstrations are also very useful for improving the learning process, as well as for adapting to different tasks when new unknown actions need to be introduced.

The problem of integrating demonstrations with RL-like algorithms has been addressed previously. Meriçli et al. [18] used a teacher to improve the robot policies, where corrective demonstrations were issued whenever the robot did not perform as well as the teacher expected. Walsh et al. [19] expanded the apprenticeship protocol of Abbel and Ng [20] by proposing a system where the teacher reviews the actions selected, and a demonstration of the optimal action is performed whenever they are not optimal. TAMER [21] is a framework where the teacher provides reinforcement rewards that evaluate the performance of the robot, and the system exploits its current model by choosing the actions that are expected to be the most highly reinforced. In these approaches, the teacher has to intervene to improve the robot behavior whenever it is not sufficiently satisfactory. By contrast, our algorithm actively requests demonstrations from the teacher whenever help is needed, thereby releasing the teacher from having to monitor the system continuously.

Active demonstration requests have been included in algorithms with confidence thresholds [22], which request demonstrations for a specific part of the state space whenever the system is not sure about the expected behavior. A confidence-based method was also described in [23], which was combined with supplementary corrective demonstrations in error cases. Agostini et al. [24] request demonstrations from the teacher when the planner cannot find a solution with its current set of rules. Our approach combines active demonstration requests with autonomous exploration. Because the teacher's time is considered to be very valuable, demonstration requests should be limited and replaced with exploration whenever possible.

When a demonstration is requested from the teacher, he does not know which parts of the model are already known. In many tasks, several actions may be selected at a given time to complete the task. However, if no guidance is provided to the teacher, he may demonstrate an action that is already known by the system. In the model, the actions are represented as a set of action rules, which can be applied over a state space. Therefore, we propose the use of a rule analysis approach to provide some guidance to the teacher so he can demonstrate the unknown parts needed by the decision maker. To explain failures when planning such models, Göbelbecker et al. [25] designed a method for finding "excuses", which are changes made to the state that make the planner find a solution. Based on these excuses, we analyze the rules to provide guidance to the teacher. Moreover, because we use a greedy learning algorithm, excuses are also useful for finding alternative models that explain unexpected states.

In summary, we propose a RL algorithm that can request demonstrations from a teacher whenever it requires new unknown actions to complete a task, which significantly reduces the number of experiences required to learn. This approach