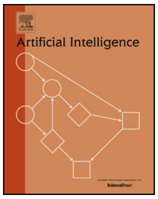
Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint



Intrinsically motivated model learning for developing curious robots

Todd Hester [*,1], Peter Stone

Department of Computer Science, The University of Texas at Austin, United States

ARTICLE INFO

Article history:
Received in revised form 28 January 2015
Accepted 11 May 2015
Available online xxxx

Keywords:
Reinforcement learning
Exploration
Intrinsic motivation
Developmental learning
Robots

ABSTRACT

Reinforcement Learning (RL) agents are typically deployed to learn a specific, concrete task based on a pre-defined reward function. However, in some cases an agent may be able to gain experience in the domain prior to being given a task. In such cases, intrinsic motivation can be used to enable the agent to learn a useful model of the environment that is likely to help it learn its eventual tasks more efficiently. This paradigm fits robots particularly well, as they need to learn about their own dynamics and affordances which can be applied to many different tasks. This article presents the TEXPLORE with Variance-And-Novelty-Intrinsic-Rewards algorithm (TEXPLORE-VANIR), an intrinsically motivated model-based RL algorithm. The algorithm learns models of the transition dynamics of a domain using random forests. It calculates two different intrinsic motivations from this model: one to explore where the model is uncertain, and one to acquire novel experiences that the model has not yet been trained on. This article presents experiments demonstrating that the combination of these two intrinsic rewards enables the algorithm to learn an accurate model of a domain with no external rewards and that the learned model can be used afterward to perform tasks in the domain. While learning the model, the agent explores the domain in a developing and curious way, progressively learning more complex skills. In addition, the experiments show that combining the agent's intrinsic rewards with external task rewards enables the agent to learn faster than using external rewards alone. We also present results demonstrating the applicability of this approach to learning on robots.

1. Introduction

Reinforcement Learning (RL) agents could be useful in society because of their ability to learn and adapt to new environments and tasks. Traditionally, RL agents learn to accomplish a specific, concrete task based on a pre-defined reward function. However, in some cases an agent may be able to gain experience in the domain prior to being given this task. Learning in this way is particularly useful for robots, as they must learn about their own dynamics and their environment before learning how to perform specific tasks. For example, a future domestic robot may be placed in a home and only later given various tasks to accomplish. In such cases, intrinsic motivation can be used to enable the agent to learn a useful model of its dynamics and the environment that can help it learn its eventual tasks more efficiently.

[*] Corresponding author.
E-mail address: todd.hester@gmail.com (T. Hester).

[1] Todd Hester is currently at Nest Labs in Palo Alto, CA.

http://dx.doi.org/10.1016/j.artint.2015.05.002
0004-3702/© 2015 Elsevier B.V. All rights reserved.

Please cite this article in press as: T. Hester, P. Stone, Intrinsically motivated model learning for developing curious robots, Artificial Intelligence (2015), http://dx.doi.org/10.1016/j.artint.2015.05.002

Past work on intrinsically motivated agents arises from two different goals [1]. The first goal comes from the active learning community, which uses intrinsic motivation to improve the sample efficiency of RL. Their goal is to help the agent to maximize its knowledge about the world and its ability to control it. The second goal comes from the developmental learning community, and is to enable cumulative, open-ended learning on robots. Our goal is to use intrinsic motivation towards both goals: 1) to improve the sample efficiency of learning, particularly in tasks with little or no external rewards; and 2) to enable the agent to perform open-ended learning without external rewards.

This article presents an intrinsically motivated model-based RL algorithm, called TEXPLORE with Variance-And-Novelty-Intrinsic-Rewards (TEXPLORE-VANIR), that uses intrinsic motivation both for improved sample efficiency and to give the agent a curiosity drive. The agent is based on a model-based RL framework and is motivated to learn models of domains without external rewards as efficiently as possible. TEXPLORE-VANIR combines model learning through the use of random forests with two unique intrinsic rewards calculated from this model. The first reward is based on *variance* in its models' predictions to drive the agent to explore where its model is uncertain. The second reward drives the agent to *novel* states which are the most different from what its models have been trained on. The combination of these two rewards enables the agent to explore in a developing curious way, learn progressively more complex skills, and learn a useful model of the domain very efficiently.

This article presents three main contributions:

1. Novel methods for obtaining intrinsic rewards from a random-forest-based model of the world.
2. The TEXPLORE-VANIR algorithm for intrinsically motivated model learning, which has been released open-source as an ROS package: http://www.ros.org/wiki/rl-texplore-ros-pkg.
3. Empirical evaluations of the algorithm, both in a simulated domain and on a physical robot.

Section 2 presents background on reinforcement learning and Markov Decision Processes. Section 3 presents work related to TEXPLORE-VANIR in the areas of reinforcement learning and intrinsic motivation. Section 4 presents the TEXPLORE-VANIR algorithm, including its approach to model learning and how its intrinsic rewards are calculated. Section 5 presents experiments showing that TEXPLORE-VANIR: 1) learns a model more efficiently than other methods; 2) explores in a developing, curious way; and 3) can use its learned model later to perform tasks specified by a reward function. In addition, it shows that the agent can use the intrinsic rewards in conjunction with external rewards to learn a task faster than if using external rewards alone. Section 6 presents details on the code release of TEXPLORE-VANIR. Finally, Section 7 concludes the paper.

2. Background

This section presents background on Reinforcement Learning (RL). We adopt the standard Markov Decision Process (MDP) formalism for this work [2]. An MDP is defined by a tuple $\langle S, A, R, T\rangle$, which consists of a set of states S, a set of actions A, a reward function $R(s,a)$, and a transition function $T(s,a,s') = P(s'|s,a)$. In each state $s \in S$, the agent takes an action $a \in A$. Upon taking this action, the agent receives a reward $R(s,a)$ and reaches a new state s', determined from the probability distribution $P(s'|s,a)$. Many domains utilize a factored state representation, where the state s is represented by a vector of n state variables: $s = \langle x_1, x_2, \ldots, x_n\rangle$. A policy π specifies for each state which action the agent will take.

The goal of the agent is to find the policy π mapping states to actions that maximizes the expected discounted total reward over the agent's lifetime. The value $Q^{\pi}(s,a)$ of a given state-action pair (s,a) is an estimate of the expected future reward that can be obtained from (s,a) when following policy π. The optimal value function $Q^*(s,a)$ provides maximal values in all states and is determined by solving the Bellman equation:

$$Q^*(s,a) = R(s,a) + \gamma \sum_{s'} P(s'|s,a) \max_{a'} Q^*(s',a'), \tag{1}$$

where $0 < \gamma < 1$ is the discount factor. The optimal policy π^* is then:

$$\pi^*(s) = \operatorname{argmax}_a Q^*(s,a). \tag{2}$$

RL methods fall into two general classes: model-based and model-free methods. Model-based RL methods learn a model of the domain by approximating $R(s,a)$ and $P(s'|s,a)$ for each state and action. The agent can then calculate a policy (i.e. plan) using this model. Model-free methods update the values of actions only when taking them in the real task. One of the advantages of model-based methods is their ability to plan multi-step exploration trajectories. The agent can plan a policy to reach intrinsic rewards added into its model to drive exploration to interesting state-actions.

This work takes the approach of using a model-based RL algorithm in a domain with no external rewards. This approach can be thought of as a pure exploration problem, where the agent's goal is simply to learn as much about the world as possible. TEXPLORE-VANIR extends a model-based RL algorithm called TEXPLORE [3] to use intrinsic motivation to quickly learn an accurate model in domains with no external rewards.

Please cite this article in press as: T. Hester, P. Stone, Intrinsically motivated model learning for developing curious robots, Artificial Intelligence (2015), http://dx.doi.org/10.1016/j.artint.2015.05.002