Automatic Induction of Generalization Hierarchies for Reinforcement Learning

Nicholas K. Jong
The University of Texas at Austin
1 University Station C0500
Austin, Texas 78712-0233
nkj@cs.utexas.edu

1. INTRODUCTION

In popular fiction, artificially intelligent agents experience the same world as humans. A spacecraft computer in *2001: A Space Odyssey* can attempt to stop the astronauts who want to deactivate it. A droid in *Star Wars* can infiltrate a space station and help rescue a princess. A robot in *Short Circuit* can discover the value of life and reject its military programming. These machines have the capacity to behave effectively in novel situations that their creators could not have explicitly anticipated.

In reality, current *autonomous agents* lack the ability to adapt to the complex structure of the human world. To achieve acceptable performance, their creators must instill either an engineered behavior policy or an ability to formulate and execute plans automatically. For this latter purpose, human designers must formalize domain knowledge that accurately captures the dynamics of the world for all those situations the agent will encounter, and they must design a representation for this knowledge that accomodates the agent's reasoning ability. The human effort required limits this approach to very constrained tasks. A key difficulty is robustness to uncertainty, due both to limitations in human knowledge and to inherent stochasticity in the world.

Research into reinforcement learning (RL) provides an attractive alternative paradigm for controlling autonomous agents. In this framework, *learning agents* rely primarily on experience data to compute behavior policies that attempt to maximize rewards over time. One broad class of RL algorithms learns the agent's optimal state-action value function, which estimates the cumulative reward possible as a function of a proposed action and the state of the world. Behaving greedily with respect to the optimal value function yields optimal behavior. Early theoretical results showed that even very simple algorithms could converge in the limit to the optimal value function for any finite problem,[1] even with no prior knowledge regarding the effects of actions or the goal of the task.

In practice, a shift from human-supplied knowledge to machine learning trades one limit on scalability for another. The theory underlying most RL algorithms assumes that every distinct situation may be arbitrarily different from every other situation. The canonical theoretical results therefore rely on visiting every distinct state infinitely often! Because the real world contains an infinite number of states, an agent cannot rely on visiting the same state more than once. When confronted with a novel situation, it must generalize when possible from past experiences in similar situations. Otherwise, the agent may spend its entire lifetime taking purely exploratory actions, never returning to a previously visited state to exploit what it learns! Insufficient generalization thus implies excessive learning time in real-world problems. At the other extreme, an agent believing that all situations are the same would always execute the same action. Excessive generalization thus prevents a learning agent from attaining intelligent behavior.

The correct generalization scheme is therefore essential to effective learning in real-world problems. However, most RL algorithms to date rely on a human designer to choose the representations that determine how the agent generalizes. This thesis advocates *scientific agents* that discover for themselves qualitative structure allowing them to generalize what they learn. Such agents would be scientific in the sense of allowing experience to revise broad hypotheses that enable more effective control over the environment. The next section describes the ideas that underpin the proposed development of such an agent, which will address the following question:

> **Can the automatic search for generalization schemes allow a reinforcement-learning agent to discover the domain knowledge necessary for coping with sophisticated, real-world problems?**

2. A LANGUAGE OF GENERALIZATION

A scientific agent must have a vocabulary with which to build hypotheses about its environment. This vocabulary must simultaneously be expressive enough to represent useful real-world knowledge and simple enough for an autonomous agent to apply effectively. This thesis will develop an agent that employs a vocabulary consisting of hierarchies of abstract models. To illustrate these concepts, consider the human example of a man considering how to commute to work on a particular day. From prior experience, the man has a mental model allowing him to predict the consequences of driving his car to work, including the amount of time the drive would consume. This model is abstract in the sense that it depends on a small subset of all the factors the man could consider. The time required does not depend

[1]The problem must allow the agent to visit every state infinitely often.

Cite as: Extended Thesis Abstract for Doctoral Mentoring Program, N. K. Jong, Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008), Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, pp. 1740-1741.

on what the man puts in the trunk or what he needs to do once he arrives at the office, but it may depend on variables such as whether the commute would take place during rush hour. The model is also abstract in the sense that it is hierarchical: taking the car to work is an abstract action that actually consists of a sequence of lower-level actions, each of which may recursively have its own abstract model. Especially at the lowest levels of this hierarchy, the models must employ approximation techniques to cope with continuous state spaces, which are natural in real-world problems.

The preceding example reveals four integral concepts. Previous research has developed each of these four concepts in the context of RL:

Function approximation is the representation of an in-principle-arbitrary continuous value function using a member of a restricted family of functions, typically defined by a finite parameterization.

Model estimation is an approach to RL that explicitly learns the effect of each action in each state, from which the optimal value function is computed.

State abstraction is a technique that assumes that certain states are completely equivalent, allowing an agent to reason in a smaller, abstract state space.

Temporal abstraction is a technique that selects actions by recursively decomposing temporally extended abstract actions into sequences of shorter actions, allowing an agent to reason at an abstract time scale.

Comparatively little work examines the compatibility and synergies among these techniques, although each has been the focus of a substantial body of research. This thesis will explicitly examine how to combine these techniques to form the core of a scientific agent. The synthesis of these ideas must allow such an agent to discover representation autonomously, in contrast to the bulk of research into abstraction, in which human users must supply the abstractions manually.

To cope with an unknown, continuous environment in the absence of prior knowledge, the agent will rely at its foundation on instance-based function approximation. Instance-based approaches only allow generalization in local neighborhoods, permitting the agent to generalize conservatively in the early stages of learning. To explore the environment as efficiently as possible, the agent will combine function approximation with model-based RL. Instance-based models will both permit efficient reuse of existing data, and reasoning about uncertainty in the model can direct the agent towards the most informative new data.

Instance-based approximation suffers from a major drawback: it does not scale well to high-dimensional problems, which are common in the real world. The scientific agent will introduce abstractions specifically to address this problem. One simple but practical family of state abstractions simply ignores some subset of the state features, reducing the dimensionality of the problem. This approach is inspired by the observation that many real-world problems have structure implicit in the representation of the state as a feature vector. In many cases, the one-step model, value function, or optimal policy at a given state may be independent of certain elements of this vector, and a scientific agent can benefit by noticing such structure.

One conjecture of this thesis is that the state-dependent set of relevant features provides a valuable basis for defining temporal abstractions. A sequence of actions executed using a given set of state features can be construed as a single temporally abstract action, whose execution lasts as long as the agent employs that state abstraction. In fact, the definition of such temporal abstractions seems necessary to apply discovered state abstractions safely, since the naive application of state abstractions can corrupt the learned value function. A scientific agent can create abstract actions that activate the state abstractions it has discovered, allowing it to learn when to apply state abstractions in the same way that any reinforcement learning agent learns when to apply actions.

The online discovery and implementation of abstractions thus allows the scientific agent to grow beyond its initial conversative generalization scheme. As the agent gathers more data, it adds more abstract actions to the reinforcement learning problem it is solving. As the set of actions evolves, the one-step model, value function, and set of optimal policies also change, perhaps supporting the discovery of new higher-level abstractions. The growing hierarchy of abstractions should allow the agent to develop increasingly sophisticated behaviors and to cope with increasingly complex environments.

3. CONTRIBUTIONS

The initial contributions of this thesis synthesized previously disparate lines of RL research. These contributions include Fitted R-MAX, an algorithm that combines the function approximation of fitted value iteration with the model-based exploration of R-MAX; R-MAXQ, an algorithm that combines the theoretical guanratees of R-MAX with the hierarchical decomposition of MAXQ; and some preliminary results on discovering partial state abstractions that can be encapsulated in temporal abstractions to speed up learning (for a model-free RL algorithm without function approximation).

This thesis has also contributed new perspectives on the issue of generalization, which underlies the unification of research into abstractions, models, and function approximation. It emphasizes how the problem of generalization properly extends beyond learning the value function to learning domain models. In particular, reasoning about generalization in models is more straightforward than reasoning about generalization in value functions. The thesis also addresses the problem of exploration, which is typically studied in finite domains where generalization is unnecessary. It may provide some insight on potential extensions to prior research on exploration, to relax the unrealistic assumption of no generalization.

Finally, this thesis will contribute a complete learning agent that can reason scientifically about its environment to determine how to generalize. This work will thus serve as an entry into the new generation of RL algorithms that learn an appropriate generalization scheme. Whereas other recently developed algorithms with this motivation attempt to determine an appropriate approximation scheme for the value function alone, the scientific agent will instead construct hierarchies of abstract models that may better capture the structure of the real world.