Towards Competence in Autonomous Agents

Özgür Şimşek
Department of Computer Science
University of Massachusetts
140 Governors Drive
Amherst, MA 01003-9264
ozgur@cs.umass.edu

Thesis Goal

My thesis aims to contribute towards building autonomous agents that are able to develop what White (1959) called *competency* over their environment—agents that are able to achieve mastery over their domain and are able to solve new problems as they arise using the knowledge and skills they acquired in the past. While the field of machine learning has made much progress in solving individual, isolated problems, progress has been slow in developing agents that are able to interact effectively with their environment and flexibly deal with new tasks. There is still a large gap between the abilities of humans in this respect and the current capabilities of autonomous agents.

My thesis will propose a number of methods for building competence in autonomous agents using the reinforcement learning (RL) framework, a computational approach to learning from interaction that has proved effective in certain types of problems (Sutton & Barto 1998). I expect that the methods I propose will extend the capabilities of RL agents in ways that are more than incremental, essentially allowing an autonomous agent to operate at a qualitatively different level.

Background

In the RL framework, the interaction between an agent and its environment takes the following form: The agent senses the state of the environment, performs one of the actions that are available in this state, and receives a numerical reward signal. The agent's objective is to maximize the expected total reward it will receive in the future.

Recent methods allow the agent to perform higher-level actions, which we will call *skills*, that are closed-loop policies over lower-level actions (Parr & Russell 1997; Sutton, Precup, & Singh 1999; Precup 2000; Dietterich 2000). An example of a skill that people use is driving. Once people learn how to drive, they no longer think in terms of the lower-level behaviors that are involved in driving. They simply choose between, for instance, driving and walking to work. In fact, most of the time, people choose among a small set of higher-level actions, which simplify their lives dramatically.

As it is with humans, the ability to automatically construct useful skills is an invaluable asset in an autonomous agent, one that can be used towards building competence (Barto, Singh, & Chentanez 2004). Designing such agents, however, remains a challenge, and the main difficulty lies in determining what constitutes a useful skill.

Proposed Research

My thesis will propose two novel and task-independent definitions of a useful skill, develop algorithms for autonomous agents to efficiently acquire such skills, and evaluate the effectiveness of these skills towards competence in agents.

The first class of skills I propose uses the notion of an *access state*, a state that allows the agent to transition to a part of the state space that is otherwise unavailable or difficult to reach from its current region. A simple navigational example of an access state is a doorway between two rooms. These states are closely related to several types of subgoals previously proposed in the RL literature (McGovern & Barto 2001; Menache, Mannor, & Shimkin 2002; Mannor *et al.* 2004).

I hypothesize that behaviors that take the agent to the access states of the environment are useful skills. First, easy access to these states allows more efficient exploration of the state space by providing direct access to those regions that are difficult to reach. Thus, these skills are useful in solving a single, isolated task. But more importantly, these are skills that are useful in a variety of problems in the same domain—getting to the doorway is useful regardless of what the agent needs to do in the other room.

While navigational domains are rich sources of access states—and are useful in conveying their intuitive appeal—access states capture a certain type of connectivity structure (of the state-transition graph) that exists in a broader class of problems. For example, completion of a subtask in a sequential task is an access state; so is building a tool that makes possible a new set of activities for the agent.

With respect to access states, I aim to address the following questions: Are behaviours that take the agent to access states useful towards building competence? Is it possible to quantify their utility? How can the agent efficiently identify access states and learn behaviours that take it to these states? Are access states common in real-world problems?

The second class of skills I propose are behaviors that efficiently uncover the key causal relationships in the environment—closed-loop policies for conducting experiments and interventions to understand the influential variables in the domain and their effects through explicit hypothesis formation and testing. Through the use of such skills, the agent would build a task-independent causal model of the environment, one that would be useful in solving many problems in the same domain. It is important to note here that the generated model would not include all of the variables and relationships in the domain—it would include only a subset of them, but a subset that consists of the relationships that are the strongest and the most influential.

The key questions I aim to address with respect to this second class of skills are the following: What types of variables should the agent include in its model, in other words, what constitutes an influential variable? How can an agent efficiently identify such variables and uncover the causal relationships among them? Is building such a causal model of the environment useful towards building competence? Can we quantify its utility?

Progress to Date

Most of my research to date has focused on the class of skills that takes the agent to access states of the environment. I have published two papers, with my advisor A. G. Barto and colleague A. P. Wolfe, demonstrating the utility of access states as subgoals in a number of tasks and proposing two methods for identifying them efficiently (Şimşek & Barto 2004; Şimşek, Wolfe, & Barto 2004).

The first method, the relative novelty algorithm (RN), is based on the observation that access states will be more likely to introduce short-term novelty than other states in the environment. The second method, L-Cut, is based on the observation that access states will be more likely to lie between two densely-connected regions of the state-transition graph. Both of these algorithms are simple, effective, and have low computational complexity. I am currently evaluating the following hypothesis: The utility of access states lies in the connectivity structure of the state-transition graph; more specifically, it is correlated with the graph-theoretic measure *betweenness centrality*, the proportion of shortest paths on the graph that pass through a given node. Furthermore, the success of RN and L-Cut lies in their ability to provide simple but accurate estimates of this measure.

Ongoing work with respect to access states include building a repository of domains, borrowing from the literature and creating new ones, to explore whether the *access state* concept is useful in general towards building competence in agents.

My work on the second class of skills I propose is in its early stages. I am currently formulating ideas on components of a skill set aimed at discovering the causal structure in the domain, building on the extensive work on computational methods for building causal models in the machine learning literature, as well as the existing research on how people build causal models (Pearl 2000; Gopnik *et al.* 2004; Tenenbaum & Griffiths 2001).

References

Barto, A. G.; Singh, S.; and Chentanez, N. 2004. Intrinsically motivated learning of hierarchical collections of skills. In *Proceedings of the Third International Conference on Developmental Learning*.

Dietterich, T. G. 2000. Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of Artificial Intelligence Research* 13:227–303.

Gopnik, A.; Glymour, C.; Sobel, D. M.; Schultz, L. E.; and Kushnir, T. 2004. A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review* 111:3–32.

Mannor, S.; Menache, I.; Hoze, A.; and Klein, U. 2004. Dynamic abstraction in reinforcement learning via clustering. In *Proceedings of the Twenty-First International Conference on Machine Learning*.

McGovern, A. E., and Barto, A. G. 2001. Automatic discovery of subgoals in reinforcement learning using diverse density. In *Proceedings of the Eighteenth International Conference on Machine Learning*, 361–368.

Menache, I.; Mannor, S.; and Shimkin, N. 2002. Q-cut — dynamic discovery of sub-goals in reinforcement learning. In *Proceedings of the European Conference on Machine Learning*, 295–306.

Parr, R., and Russell, S. 1997. Reinforcement learning with hierarchies of machines. In Jordan, M. I.; Kearns, M. J.; and Solla, S. A., eds., *Advances in Neural Information Processing Systems*, volume 10. The MIT Press.

Pearl, J. 2000. *Causality : Models, Reasoning, and Inference*. New York, NY: Cambridge University Press.

Precup, D. 2000. *Temporal abstraction in reinforcement learning*. Ph.D. Dissertation, University of Massachusetts Amherst.

Şimşek, Ö., and Barto, A. G. 2004. Using relative novelty to identify useful temporal abstractions in reinforcement learning. In *Proceedings of the Twenty-First International Conference on Machine Learning*.

Şimşek, Ö.; Wolfe, A. P.; and Barto, A. G. 2004. Local graph partitioning as a basis for generating temporally-extended actions in reinforcement learning. In *Proceedings of the AAAI-04 Workshop on Learning and Planning in Markov Processes - Advances and Challenges*.

Sutton, R. S., and Barto, A. G. 1998. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Sutton, R. S.; Precup, D.; and Singh, S. P. 1999. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence* 112(1-2):181–211.

Tenenbaum, J. B., and Griffiths, T. L. 2001. Structure learning in human causal induction. In *Advances in Neural Information Processing Systems*, volume 13, 59–65. The MIT Press.

White, R. W. 1959. Motivation reconsidered: The concept of competence. *Psychological Review* 66:297–333.