A Motivational System that Drives the Development of Activity

Matthew D. Schmill
Computer Science Department
University of Massachusetts
Amherst, MA 01002
schmill@cs.umass.edu

Paul R. Cohen
Computer Science Department
University of Massachusetts
Amherst, MA 01002
cohen@cs.umass.edu

Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning; I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search

General Terms

Algorithms

1. INTRODUCTION

When is an agent truly autonomous? Steels distinguishes *autonomous* systems from *automatic* systems in the following way:

> AI systems built using the classical approach are not autonomous, although they are automatic. Knowledge has been extracted from experts and put into the system explicitly [...] The resulting systems can solve an infinite set of problems [...] but these systems can never step outside the boundaries of what was foreseen by the designers because they cannot change their own behavior in a fundamental way. [7]

Related work by Luck & d'Inverno [3], Norman & Long [4] and others forward the ideas of *self-generation* (of goals) and *motivation* as a step towards autonomy by allowing an agent to select the goals it will pursue dynamically, under the control of its own motivational system. We contend that true autonomy, in the sense described by Steels, can only be achieved by a system that develops its own knowledge, and creates its own goals and behaviors grounded in this knowledge.

In this end, we have developed a theory of the development of activity in intelligent, autonomous agents that comprises three complementary processes. In the first process, an agent generates its own goals. Rather than choosing among arbitrary or exogenously-given states of the world, the agent generates goals of the form "engage in activity A" by a process we call *planning to act*. It then decides among available goals by ranking them according to a motivational system that takes into account the needs of the agent. In the second process, the agent uses a means-ends analysis (MEA) planner to build plans to achieve its goals. Plans that prove successful can be added to the agent's library of *activities* and retrieved for later use or revision. Finally, in the third process, our system learns from its experience to produce and refine operator models that inform the motivational system and planner. Guided by the motivational system, our agents learn new activities and their competence grows gradually to encompass all the activity an environment affords. Our approach parallels Piaget's developmental psychology of infants, in which executing and extending schemas (i.e., executing plans and extending them to achieve their preconditions) is rewarding (i.e., preferred by the motivational system) [2].

Due to space constraints, we limit the discussion here to goal generation and selection via the motivational system, and a short discussion of some results working with the system in simulated domains. More detail on the planner and modeling systems can be found in [5] and full details of the experimental results can be found in [6].

2. GOAL GENERATION AND SELECTION

The problem of how to generate goals for a classical planner to achieve is one that is traditionally not automated. Planning goals are, in general, specified as desirable sensory or perceptual states by some exogenous source, typically the experimenter. An account of development, though, must explain how an agent produces its own goals.

For a reasonably complex agent, such as a Pioneer-2 mobile robot, with 60 real-valued sensors, the space of possible sensory states is huge, and large regions of that space are unreachable or practically indistinguishable from each other. The question of how to generate goals is under-constrained.

A simple observation on how human children spend their time during development provides leverage on this problem. Piaget noted that children seem to spend their time exercising *schemas*, or simple sensorimotor routines. The goals of an agent, and those things that are rewarding to an agent, seem to be *activities*, not sensory configurations. If we adopt a scheme where the goals of an agent are activities, not sensory states, we limit the goal space to a finite space of achievable goals. We call this philosophy of goal generation *planning to act*.

With the space of goals restricted to activities that the

AAMAS'02, July 15-19, 2002, Bologna, Italy.

agent knows about, goal selection can be accomplished by the use of a motivational system similar to those described by Norman [4] and others. A motivational system defines a preference relation $\psi(a_i, a_j, s_t)$ that states a preference to engage in activity a_i over a_j in state s_t.

Our motivational system comprises a set of separate *motivational factors*, which correspond to basic drives of an agent, like hunger, fatigue, and curiosity. Each motivational factor is represented numerically by a coefficient μ_f, which is genetic, and an expected change E_Δ, which is learned. The *desirability* of any activity a_n in state s_t is defined as

$$d(a_n, s_t) = \sum_{F \in \mathcal{F}} \mu_F(s_t) E_\Delta(a_n, s_t) \tag{1}$$

and the preference relation simply compares the desirability of any two activities.

3. EVALUATION

There are two primary roles for the motivational system to play in our system: it must *exploit* existing activities to keep the agent functioning properly, and it must *explore* in order to learn new activities that may provide better opportunities for exploitation in the future. These roles are at odds, and managing this so-called exploitation versus exploration tradeoff is a well studied problem in reinforcement learning [1].

Our system manages this tradeoff my allowing the importance of different motivational factors to rise and fall, at times allowing the curiosity factor to dominate, while at other times, allowing vegetative needs to take control of the robots actions through ψ. The agent discovers new activities by generating and executing plans based on incomplete operator models.

To demonstrate that our motivational system manages the tradeoff between exploitation and exploration successfully, we implemented two simulated domains. The first is a simulated factory domain, in which a robot with a paint gun must discover that it can achieve rewards that satisfy its motivations by picking up unpainted blocks and painting them. In order to achieve rewards, it must avoid overheating through periodic resting, and it must keep its paint gun loaded with paint cartridges. The second domain is implemented by randomly generating Markov decision processes (MDPs). In this domain, rewards are randomly distributed among the MDP, and it is the job of the agent to find the rewarding activities and exploit them.

We ran trials with our system on the simulated painting robot, a domain with 5 actions and a total 19 possible outcomes of those 5 actions, on a smaller MDP with 5 actions, 10 states, and 22 outcomes, and a larger MDP with 10 actions, 20 states, and 44 outcomes. In the simulated robot domain, the robot had discovered all 19 outcomes just 60 steps into the simulation. By the time the robot had taken 100 simulated actions, the effect of curiosity had faded to the point where the robot settled into a policy of painting blocks, refilling its paint gun, and resting.

Results similar to those found with the simulated robot were generated in our trials with the randomly generated MDPs. In the smaller process, the agent had discovered all 22 outcomes by step 100 into the simulation, and curiosity had faded after 200. The agent settled into a sequence of two plans, each of length 3, which maximized its rewards in the randomly generated domain. In the larger MDP, the agent discovered a rewarding activity almost immediately. While this influenced its early behavior, the agent still managed to discover 39 of the 44 outcomes in 140 steps. This, we believe, is due to the agent attempting to build plans for the rewarding activity with poor models. The agent builds plans that are destined to fail, but in so doing, the agent discovers new outcomes. In so doing, at around step 140, the agent happened upon a second rewarding outcome that formed a cycle with the first rewarding outcome. By 200 steps, the agent was already showing signs of settling in on this rewarding policy where one rewarding activity led to a second.

4. CONCLUSIONS

We have developed a theory of the development of activity in autonomous agents. This theory is based on means-ends analysis planning because of its developmental plausibility and the declarative, compositional nature of its representations, which we believe will useful in the development of related types of conceptual structure such as classes and language.

The course of development in our system is determined by automated goal generation and selection. Planning to act is a process by which an agent plans to engage in activities, rather than achieve world states. This allows the agent to limit its attention to a small, fixed set of achievable goals that it can evaluate according to its motivational system. The motivational system we use is based on a combination of motivational factors like hunger and curiosity, which are modeled numerically, and change over time.

We have tested our system on simulated domains and shown that it effectively manages the tradeoff between exploration and exploitation. In our current work, we have transitioned this system to the Pioneer-2 mobile robot, where we are optimistic that a rich set of activities will emerge as the robot explores its environment.

5. REFERENCES

[1] R. Dearden, N. Friedman, and S. Russell. Bayesian q-learning. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, Madison, WI, 1998.

[2] H. P. Ginsberg and S. Opper. *Piaget's Theory of Intellectual Development*. Prentice Hall, Englewood Cliffs, NJ, 1988. 3rd Edition.

[3] M. Luck and M. d'Inverno. Motivated behavior for goal adoption. In Zhang and Lukose, editors, *Multi-Agent Systems: Theories, Languages and Applications - Proceedings of the Fourth Australian Workshop on Distributed Artificial Intelligence*, pages 58–73. Springer-Verlag: Heidelberg, Germany, 1998.

[4] T. J. Norman and D. Long. Goal creation in motivated agents. In M. Wooldridge and N. R. Jennings, editors, *Intelligent Agents: Theories, Architectures, and Languages (LNAI Volume 890)*, pages 277–290. Springer-Verlag: Heidelberg, Germany, 1995.

[5] M. Schmill, T. Oates, and P. R. Cohen. Learning planning operators in real-world, partially observable environments. In *Proceedings Fifth International Conference on Artificial Planning and Scheduling*, pages 246–253. AAAI Press, 2000.

[6] M. D. Schmill and P. Cohen. A motivational system that drives the development of activity. Technical Report TR 01-12, Computer Science Department, University of Massachusetts, Amherst, MA, 2001.

[7] L. Steels. When are robots intelligent autonomous agents? *Journal of Robotics and Autonomous Systems*, 15:3–9, 1995.