Motivated Agents

Kathryn Kasmarik[1], William Uther, Mary-Lou Maher[1]
National ICT Australia[*], [1]University of Sydney
kkas0686@it.usyd.edu.au, william.uther@nicta.com.au, mary@arch.it.usyd.edu.au

1 Introduction

This poster presents a model for motivated agents governed by an internal, domain independent motivation process rather than domain specific rewards, goals, or examples provided by an external teacher. Internal motivation is desirable in complex, dynamic environments where hard-coded domain theories can only approximate the true state of the world and where pre-programmed goals can become obsolete.

Early work with motivated agents focused on the use of domain specific motives [Sloman and Croucher, 1981] or environment modelling by focusing attention on situations with the highest potential for learning [Kaplan and Oudeyer, 2004]. More recent work with intrinsically motivated reinforcement learning agents [Singh et al., 2005] has produced agents that, rather than learning a model, learn new behavioural *options* [Precup et al., 1998]. An option or simply a *behaviour* is a whole course of action that achieves some sub-goal. Our model extends previous work by defining general structures for events, attention focus and motivation.

2 The Motivated Agent Model

Our model for a motivated agent is shown in Figure 1. The agent has three types of structures: sensors, memory and effectors. These structures are connected by three processes: sensation, motivation and action. The agent functions in a continuous sensation-motivation-action loop.

Sensors receive raw data in the form of an n-tuple of state variables $\{x_1, x_2, x_3, \ldots x_n\}$ describing the current state of the agent and its environment. The n-tuple has a particular format. Variables x_1 to x_k comprise data about the state of the agent. Variables x_{k+1} to x_n comprise data about other objects sensed by the agent. This includes absolute data as well as relative data such as the location of an object relative to the agent. This general format is domain independent and the inclusion of relative values makes it possible for the agent to learn general behaviours relative to itself.

Sensation transforms raw data received from sensors into structures called *events.* Events encapsulate two recognisable occurrences among state variables values: increases and decreases between one state and the next. Individual events and the current state n-tuple are incorporated into *memory.*

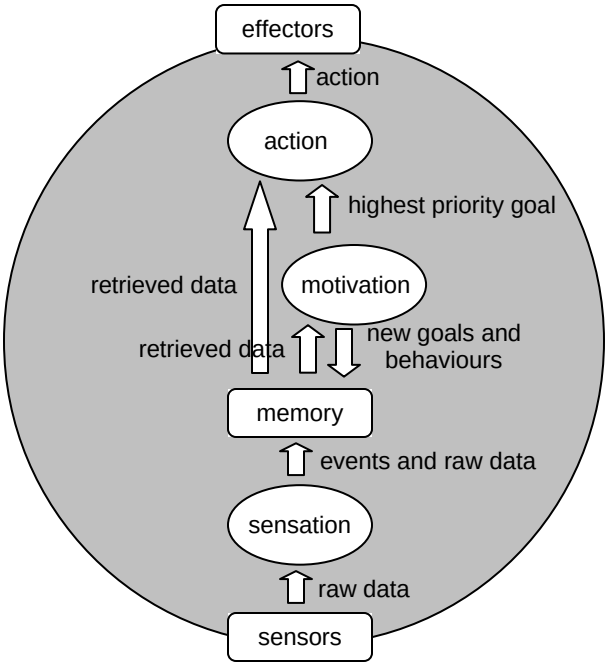

Figure 1 – The motivated agent model.

Memory is a cumulative record of events, actions, and goals. Initially, memory is empty save for a set of primitive actions that the agent can perform in the world.

Motivation motivates agents to understand and repeat interesting events that occur in their environment. We characterise interesting events as changes in the world that occur infrequently. Agents are motivated to create goals to understand and repeat interesting events by identifying situations in which they can learn how to make the event recur using a temporal difference Q-learner (TDQL) [Sutton and Barto, 2000]. Goals are *masked* so that events that occur with equal or lesser frequency than the one being pursued are ignored. Masking increases learning efficiency by focusing

[*] National ICT Australia is funded by the Australian Government's Backing Australia's Ability initiative, in part through the Australian Research Council.

attention and reducing the size of the state space. Once an agent can repeat an interesting event at will, it can encapsulate its new knowledge as a *behaviour*. Masking ensures that behaviours are independent of the situation in which they were learned. A behaviour can be reused either as a pre-planned course of action for achieving new goals similar to the one from which it was originally created or as a building block when learning to solve more complex goals.

Action reasons about which *effectors* will further the agent's progress towards its current goal.

Effectors are the means by which actions are achieved. They allow the agent to cause a direct change to the world.

3 Motivated Agents in Practice

We demonstrate our agent model using a robot guide. This domain, based on Dietterich's taxi domain [2000] and illustrated in Figure 2, contains avatars who need to be guided to particular locations. Such situations are common in large scale virtual worlds where new citizens can easily become lost. There are four possible sources and destinations for avatars. The robot has twelve effectors controlling the movements of its legs and can choose to start or stop guiding an avatar. The robot's sensors can perceive the absolute co-ordinates of the agent, the elevations of its legs, the co-ordinates of its legs relative to its body and the absolute and relative co-ordinates of an avatar and its destination.

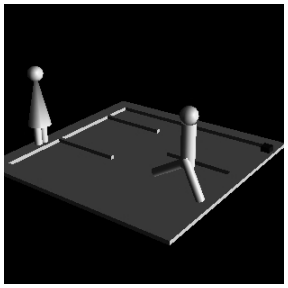

Figure 2 – A guide robot.

After a period of exploration, the robot notices that increases and decreases in its location and the location of other avatars are infrequent. The agent is motivated to create and pursue goals to repeat these changes. Over time the agent learns walking behaviours, such as Behaviour-1, to change its location. These behaviours are independent of the location in which they were learned and the position of other avatars. The agent uses these walking behaviours to develop path-following behaviours, such as Behaviour-2, to change the location of the avatar.

Behaviour-1 [lift left foot, move left foot forwards, put-down left foot, lift right foot, move right foot forwards, put-down right foot]
Behaviour-2 [guide, Behaviour-1, Behaviour-1, Behaviour-1, Behaviour-1, Behaviour-1, stop-guiding]

4 Empirical Results

We measured the performance of the robot guide by graphing the number of primitive actions taken to produce Behaviour-2 in response to some goal G. We then implemented a flat TDQL with a pre-programmed reward for achieving a similar behaviour that satisfies G. Figure 3 shows that the motivated agent learns Behaviour-2 more quickly than a flat TDQL can learn a similar behaviour. This is because the motivated agent can make use of abstract behaviours such as Behaviour-1while constructing Behaviour-2.

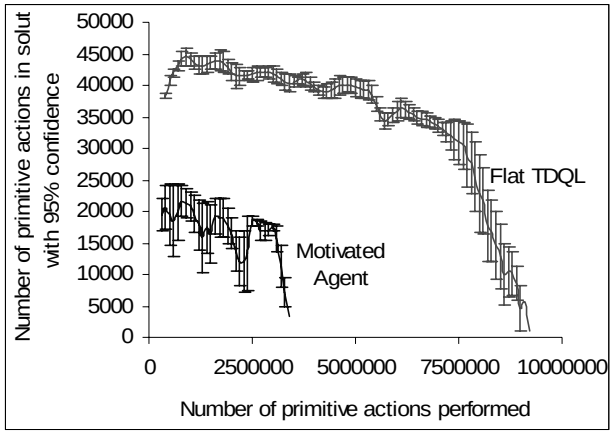

Figure 3 – Learning progress in a guide task.

5 Conclusion

Motivated agents are autonomous agents governed by internal, domain independent motivation. In addition to being able to choose their own goals they are able to learn behaviours to satisfy their goals more quickly than agents using the TDQL algorithm without motivation.

References

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

[Kaplan and Oudeyer, 2004] F. Kaplan and P-Y. Oudeyer. Intelligent adaptive curiosity: a source of self-development. In *Proceedings of the 4th International Workshop on Epigenetic Robotics*, pages 127–130, 2004.

[Precup et al., 1998] D. Precup, R. Sutton, and S. Singh. Theoretical results on reinforcement learning with temporally abstract options. In *Proceedings of the 10th European Conference on Machine Learning*, 1998.

[Singh et al., 2005] S. Singh, A. G. Barto, and N. Chentanex. Intrinsically motivated reinforcement learning. To appear in *Proceedings of Advances in Neural Information Processing Systems 17* (NIPS), 2005.

[Sloman and Croucher, 1981] A. Sloman. and M. Croucher. 1981. Why robots will have emotions, The 7th International Joint Conference on Artificial Intelligence, Vancouver, Canada, pp. 197-202.

[Sutton and Barto, 2000] R. Sutton, and A. Barto. *Reinforcement learning, an introduction*. MIT Press, 2000.