Embodied Decisions:
Models of Decision Making Within a Larger Cognitive Framework

Terrence C. Stewart (terry@ccmlab.ca)
Carleton Cognitive Modelling Lab
Institute of Cognitive Science
Carleton University

Overview

The following pages describe ongoing research within the Carleton Cognitive Modelling Lab at the Institute of Cognitive Science, Carleton University, Canada . This project is to understand and develop working models of real-world high-level domain-general non-explicit adaptive behaviour. The work draws from numerous existing computational models, organizing and situating them in such a manner than they can be directly evaluated as potential explanations as to how real creatures learn to respond to their surroundings.

To further this goal, we have developed a general framework for describing such models, identified relevant known behavioural results in real creatures, created a methodology for comparison, and specified a diverse set of models for consideration.

Rationalle

Our goal is to understand the behaviour of living creatures. To do this, we follow the standard scientific method of observing some behaviour, forming theories about the underlying mechanisms, identifying the resulting predictions, testing those predictions, using those results to form new theories, and then repeating the process again and again.

In recent years, thanks to the wide availability of computation, the theories as to how living creatures choose one action over another have become more and more complex. No longer satisfied with cognitive models involving underspecified "boxes-and-arrows" diagrams, researchers have been developing intricately specified computational models which detail the particular mechanisms whereby the creature learns about its environment, discovers the consequences of its actions, and then selects behaviours to perform. These computer simulations involve complex environments, bodies, and brains, and a wide variety of tasks to be performed. Importantly, they also support quantitative predictions, such as exactly how quickly certain sorts of tasks will be learned, or what sort of errors will be made at what rates. Much like the introduction of mathematical modelling in physics, this allows us to move away from qualitative predictions such as "people tend to choose to do things they were rewarded for doing in the past" toward being able to specify how strong this tendency is, the effects of different combinations of rewards and punishments, and other more precise predictions.

The state-of-the-art in such models is currently not at a level where we can do extensive comparisons to human performance. Instead, we will be looking at models which may be suitable for dealing with more basic, animal-level, non-explicit behaviours. This covers a vast variety of models, as most of the field of Artificial Life and a large portion of Artificial Intelligence falls under this description. A review of the existing literature reveals that "the number of architectures and working principles has grown much faster than the number of comparisons" (Guillot and Meyer 2000). Furthermore,

> Sharing work has been so difficult that researchers tend to build their own animat minds and worlds from scratch, often duplicating work that has been done elsewhere. There have been a number of attempts to re-use animat or agent minds and worlds, but the model of re-use often requires installation, or even a particular programming language.... Often, the only person who ever does experiments with an animat or agent is its author (Humphrys and O'Leary 2002).

If we are to evaluate these models in terms of their potential for explaining the functioning of real living organisms, we need to perform these comparisons. We need a way to organize these models into a common framework, allowing us to identify the key distinctions between models. As the situation is currently, fundamental features of the core processes within a particular model are confounded with peripheral features such as the structure of the creature's body, its sensory capacities, or what sorts of other cognitive abilities are introduced to support decision making. That is, each cognitive model is embedded within a larger brain, each brain within a body, each body within an environment, and each environment is set up for a particular task. Thus, when evaluating a model's performance on a particular task, we must recognize that these other aspects will also affect behaviour. If we are to compare models, we must control for these components.

Behaviours

Before discussing our approach, it is useful to identify exactly what sort of behaviours we are hoping to achieve from our models. Tasks requiring human-level introspective reasoning abilities are clearly beyond any of the models we are examining. Instead, we are going to examine these models in terms of standard behavioural tasks well-known to experimental psychology. Many of these experimental paradigms are taken from animal research, so as to ensure the lack of a requirement for an ability to explicitly reason. We have chosen to focus on a range of tasks which are widely studied in real living creatures (such as operant and classical conditioning), as this provides us with a wealth of experimental data for comparison. It has also proven to be a useful domain for comparison in previous projects of smaller scope, such as (Stewart 2000).

Our general term for these tasks is 'Embodied Decision Making'. These are situations which are embodied environmentally, in that there is a larger world around the simulated creature upon which it can act and be acted upon. In other words, the creature's actions affect its own future senses. If it decides to move toward an object, that object will appear closer in the future. This is sometimes referred to as being situated, and means that the creature is always within a tight feedback loop with its environment. The importance and impact of this has become more evident in recent years (for example, see Clark 1997). Furthermore, the decision-making process itself is vitally dependent and interconnected with the rest of the creature's mental abilities: recognizing objects, forming categories, learning from the past, interpreting feedback, and so on. Thus, we cannot study the core decision-making algorithm without also studying the effects of these various different components of the creature's mind that the decision-making system is embedded within.

The behavioural tasks we are concerned with are high-level, domain-general, non-explicit, and adaptive. High-level indicates that we are dealing with deciding between actions expressed in terms such as 'run away' or 'press the button', as opposed to low-level actions like 'contract left soleus muscle'. We are at present not concerned with the process whereby such a high-level decision is converted into a complex series of motor commands. Domain-general indicates that the cognitive processes involved are seen as being applicable to many different situations, as opposed to being specifically evolved for one purpose. Non-explicit indicates our attention to sub-human level intelligence. It should be noted that it is not clear whether human explicit reasoning is a completely separate system from non-explicit reasoning, as it does seem likely that explicit reasoning would be built on top of a pre-existing, evolutionarily prior, decision-making system. For this reason, it is quite possible that understanding non-explicit reasoning will give insights into human intelligence as well. Finally, adaptive indicates that we will be focusing on behaviours which change over time as the creature interacts with its environment over a reasonably short period of time (minutes or hours).

We have identified the following phenomena as suitable for our investigations. Each one has been widely observed in a variety of situations and species.

Acquisition: the characteristic response curve as a new stimulus is associated with a reward in operant conditioning

Extinction: the gradual decaying of response when a reward no longer occurs

Spontaneous Recovery: after an association is extinguished, a long delay in a new environment can cause it to reappear

Reinforcement Schedules: the effect of delayed and random reinforcement, different ratios, and variable time intervals

Generalization: the response to other stimuli that are similar to the one trained on

Specialization/Discrimination: learning to respond differently to two stimuli which initially give the same response

Classical Conditioning: responding to a stimulus that has been paired with a stimulus which has a 'hard-wired' response

Second Order Conditioning: responding to a stimulus that has been paired with a stimulus which has a previously learned response

Category Learning: determining if action A or B is appropriate given a particular (usually complex) stimulus

Recency Effects in Category Learning: a bias toward choosing an action that was rewarded recently (Myers 1970)

Peak Shift: training a response to stimulus A (e.g. a 1000Hz tone) but not B (e.g. a 900Hz tone) causes an even stronger response to A' (e.g. a 1050Hz tone) (Hanson 1959)

T-Maze: learning to associate a stimulus with a direction to turn in a maze at a later time

Delayed Match-To-Sample: remembering a stimulus and later identifying it from a set of stimuli

Same/Different Tasks: learning to use the concepts of 'sameness' and 'difference', versions of which should be failed by the models we are examining, as animals also fail them

This forms the set of basic abilities that we can use to compare the behaviour of our models to that of real creatures. Certainly, there are many other possibilities as well. We have chosen these based on their relatively universal results, their ease of implementation in our experimental framework, and their tendency to not be aspects which are specifically referred to in the design of the algorithms themselves.

Framework

We have developed a basic framework within which we can organize the many disparate models developed by other researchers. This allows us to examine the behavioural differences when we change only one component at a time. We can thus compare fundamentally different learning models, while keeping the rest of the virtual creature's brain, body, and environment exactly the same. Indeed, we can also keep the core learning algorithm the same, and adjust other aspects, such as the sensory capabilities, the perceptual pre-processing, the actions available, the complexity of the environment, and so on.

When we examine the adaptive models discussed in the field, there are certain common aspects which can be identified. The two core aspects are that of learning about the world, and using that knowledge to choose an action.

The first component we shall refer to as the World Model, as the system can be seen as developing its own model of the world around it. This generally takes the form of a predictive system, allowing the creature to determine the likely outcome of a particular action. Formally, this is a cognitive module which can take the current state of the world (S_t) and predict a future world state (S_{t+1}), under the influence of a particular chosen action from the creature (A_t).

The World Model systems generally learn through experience. This is done by noting that whenever the creature makes any action, the results of that action can be used to train the system. This is then a classic example of supervised learning, and a wide variety of computational models can be used for this task.

Upon examining these different World Model modules used within various researcher's models, we find two further complications. Some of these models do not do a complete prediction of the future state of the world; instead, they merely predict certain aspects of it that turn out to be useful. Others are noteable in that they are able to deal with hidden or unobserved aspects of the world (that is, they do not assume that all of the relevant aspects of the world are observable to the creature). This complicates matters in that the system needs to postulate and keep track of certain bits of knowledge, such as where some particular food was hidden, or which object was seen previously.

The second component is often designated by the term Action Selection. This is a process whereby the creature makes use of the World Model and the currently known state of the world to decide upon an action to take. To do this, various models use a variety of approaches, as befits a problem as complex as this. These methods generally include the definition of some sort of reward for the creature. The World Model can then be used to predict this reward in possible future situations. The creature thus 'plans' a sequence of future actions which may bring about that reward. Of course, since the creature cannot perfectly predict the future, and nor can it imagine all the possible actions it could take, this process of deciding upon an action must be done via a variety of heuristics. Furthermore, the various approaches need to balance exploration (trying out actions in new situations to see what happens) and exploitation (making use of previously-learned information to choose a good action). The different sorts of Action Selection modules thus differ in almost every regard, including how potential actions are thought of, how far into the future they look, and how they evaluate how 'good' an outcome will be.

While all of the models we have examined have something that can be seen as a World Model and something that can be seen as an Action Selection system, some models have aspects which are best thought of as further additions. Importantly, we can often take these additions and combine them with World Models and Action Selection modules for which they were not initially developed.

One type of extra component is the Pre-Processor. This is a system which takes the sensory inputs received by the creature and changes their representation. This change is based on it developing a new representation scheme for the given situation. The importance of this is two-fold: First, it may be able to uncover underlying similarities in the stimuli that are not evident in the initial representation. Second, it reduces the influence of potential bias caused by the researcher's decision to represent the incoming stimuli in a particular way.

This process can have a remarkable impact on the performance of the computational models. Unfortunately, this fact tends not to be recognized in many situations; researchers just choose one way of presenting data to their simulated creatures. They give their creatures a set of senses which cleanly depict the environment in terms of how the researchers themselves conceptualize the world. This leads to a worry that some results may be due to the particular representational choices made, rather than the computational model itself. Furthermore, there can be an effect in the other direction as well; the representation thought of by the researcher may be sub-optimal for the situation at hand. Results such as (Smith 2002) show the potential for a vast improvement in performance with such an approach.

Given these distinctions, we arrive at the organizational diagram for our research shown in Figure 1.

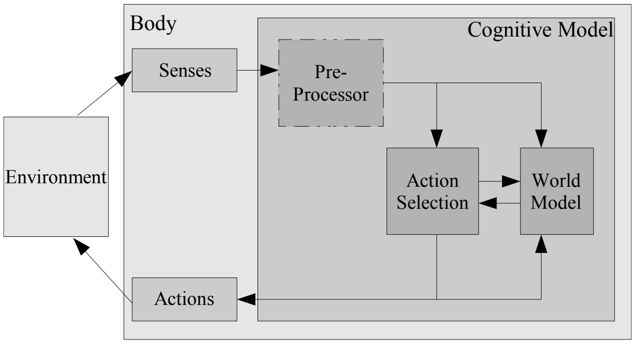

Figure 1: A generic framework for embodied decision-making models.

Many currently existing models can be interpretted as falling within this framework. Furthermore, once they are in this framework, we can mix and match components between models. For example, it is quite common to use a highly simplistic World Model, such as a look-up table (where the prediction for what will happen if the agent does

A in state S is simply based on memorizing whatever it was that happened the last time the agent was in state S and did A). However, a more complex World Model could be put in place without changing the rest of the cognitive model. This allows us to examine new models which do not exist in the literature. Most researchers restrict themselves to one (or perhaps two) possibilities for most of these components, and focus their attention on one part in particular, leaving the rest of the cognitive model with simple components. Since these various aspects of the model clearly can have an impact on the overall performance, we cannot test them in isolation.

Taking this approach allows us to examine the overall effect of having various sorts of implementations of these various modules. Some complex modules may turn out to be behaviourally indistinguishable from simpler versions of the same component. There may, indeed, be various complex interference effects between these different versions. The hope is to discover overall generalities that can lead us to determine necessary or sufficient conditions for building cognitive models which exhibit the decision-making capabilities observed in living organisms.

Comparison

Our overall task is to use the structure provided by the above diagram to allow us to rigorously investigate the performance of various computational models of basic learning. More specifically, the goal is not to determine which models (or which combinations of aspects from various models) perform the best; instead, we are looking for models which match the behaviour of real creatures. Importantly, we do not want a model which exactly matches every quirk of some particular experimental data (as doing so would lead us to over-fit models via parameter tweaking). Instead, we want models which capture the observed behaviour to some pre-specified degree of accuracy.

Our methodology for this comparison is based on equivalence testing . This is a modified version of the standard t-test, where instead of the traditional Null Hypothesis that the means of two groups are equal ($\mu_r - \mu_m = 0$), we set the Null Hypothesis to be that the difference between the means is greater than some amount ($|\mu_r - \mu_m| > \theta$). The value of θ allows us to control how tight a match we require. If we perform this statistical test, using μ_r as the real data set, and μ_m as the data from a given model, then a p-value less than 0.05 allows us to conclude with 95% certainty that the model and the real system do not differ by more than our threshold, θ. For a concise introduction to this approach, see (Streiner 2003).

Given this method for determining that the model and the real data are statistically significantly equivalent (as opposed to the traditional t-test, which allows us to find statistically significant differences), we can apply this test repeatedly (controlling for the repeated measures), resulting in a set of models which we can be reasonably sure give accurate enough predictions for the various behaviours listed above.

One further consideration is what to do with models that have parameters which affect their performance. Instead of 'tweaking' the parameters to find the model that most closely fits a given data set, we instead evaluate the model multiple times using a variety of different parameters settings. This can lead us to discover a range of parameters settings over which the model continues to perform well. This process is similar to the standard modelling practise of sensitivity analysis.

Models

The following models have been identified as suitable for this research, and have been decomposed into their constituent parts as per our framework. Once identified, each component of each model is suitable for inclusion in every other model. To make this possible, we are writing our own versions of the software for each model, based on their published descriptions. Our re-implementation allows us to replicate the original experiments and then perform our own analysis of model performance, including the effects of combining the components of the models in various ways.

Investigating each of these myriad possibilities is our continuing project. We are developing our own series of virtual test environments comparable to those used in the experimental psychology research on the behaviours listed previously. There are clearly many more models in the literature which can be examined, but we have limited ourselves to ones which use more well-known components, so as to reduce the workload required.

Model-Based Evolution (Keymeulen et al. 1998)
World Model: Look-up table
Action Selection: Genetic Algorithm
Pre-Processing: None

Reinforcement Learning (Sutton and Barto 1998)
World Model: Look-up table/Back-prop Neural Network (predicts only long-term reward, not exact future states)
Action Selection: Q-Learning, TD(λ), SARSA, and others
Pre-Processing: Manually designed to suit the task

Distributed Adaptive Control (Veogtlin and Verschure 1992)
World Model: Associationist Network
Action Selection: Contextual Control Network
Pre-Processing: None

Neural Reinforcement Learning (Touzet, 1997) (Smith, 2002)
World Model: Look-up table (long-term reward only)
Action Selection: Q-Learning
Pre-Processing: Kohonen Self-Organizing Map

Genetic Planning (Farritor and Dubowsky 2002)
World Model: Human-designed Mathematical Model
Action Selection: Genetic Algorithm
Pre-Processing: None

Recurrent Neural Networks (Tani 1998)
World Model: Recurrent Neural Networks
(can track hidden aspects of the world)
Action Selection: Association Network
Pre-Processing: Hopfield Network

Hierarchical Recurrent Neural Networks (Tani and Nolfi 1999)
World Model: Hierarchical Recurrent Neural Networks
(can track hidden aspects of the world)
Action Selection: None
Pre-Processing: None

Incremental Self-Improvement (Schmidhuber, 1997)
World Model: Look-up table (long-term reward only)
Action Selection: Curiosity-based Reinforcement
Note: biased toward making small errors in prediction
Pre-Processing: None

Skinnerbots (Touretzky and Saksida 1997) (Daw et al. 2003)
World Model: Semi-Markov Models
Note: can track hidden aspects of the world
Action Selection: Dopamine-based TD-Learning
Pre-Processing: None

It should also be noted that any supervised learning system (any system that uses previous examples of input-output pairs to determine what output to give for a new input) is potentially usable for the World Model. Furthermore, any unsupervised system (any system that takes a large quantity of input data and produces a new representation for it based on patterns found in the data) is a potential Pre-Processor.

There is clearly no consensus as to what sorts of components are better than others, or if certain components produce better models for some tasks bu not for others. Indeed, the afore-mentioned models have not even been tried on the same task as each other, much less in an environment comparable to that of a real creature.

By doing this broad-based comparison between the models and the real-world experimental results, we will achieve a deeper understanding of how these models work. Most importantly, successful models can be shown to be relevant for understanding how real cognition works. That is, we will not be concluding merely that model X performs better than model Y. Instead, we will be finding that model X is the best model we have of how embodied decision making happens in real creatures.

Certainly, there will be aspects of real behaviour that are not captured. But, once these are identified, and once we have a consistent framework for scientific investigation, we can develop further, more complete models and evaluate them. This cycle of forming hypotheses about how something works, testing those hypotheses, and developing new ones is exactly how science proceeds. We believe that our approach to structuring this research is a necessary step for us to come to understand and explain this fundamental aspect of cognition.

References

Clark, A. 1997. Being There: Putting Brain, Body and World Together Again. Massachusetts: MIT Press.

Daw, N.D., A.C. Courville, and D.S. Touretzky. 2003. Timing and partial observability in the dopamine system. In Advances in Neural Information Processing Systems 15, edited by S. Becker, S. Thrun, and K. Obermayer. Massachusetts: MIT Press.

Farritor, S. and S. Dubowsky. 2002. A Genetic Planning Method and its Application to Planetary Exploration. ASME Journal of Dynamic Systems, Measurement and Control 124 (4): 698-701.

Guillot, A. and J.A Meyer. 2000. From SAB94 to SAB2000: What's New, Animat? In From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior. Massachusetts: MIT Press.

Hanson, H. M. 1959. Effects of discrimination training on stimulus generalization. Journal of Experimental Psychology 58: 321-334.

Humphrys, M. and C. O'Leary. 2002. Constructing complex minds through multiple authors. In From Animals To Animats 7: The 7th International Conference on the Simulation of Adaptive Behavior. Massachusetts: MIT Press.

Kakade, S. and P. Dayan. 2000. Acquisition in Autoshaping. In Advances in Neural Information Processing Systems 12, edited by Solla, S., Leen, T., and Müller, K. Massachusetts: MIT Press.

Keymeulen D., M. Iwata, Y. Kuniyoshi, and T. Fliguchi. 1998. Comparison between an Off-line Model-free and an On-line Model-based Evolution applied to a Robotics Navigation System using Evolvable Hardware. In Proceedings of the 6th International Conference on Artificial Life. Massachusetts: MIT Press.

Myers, J. L. 1970. Sequential choice behavior. In The psychology of learning and motivation: Volume 4, edited by Bower, G. H. and Spence J. T. New York: Academic Press.

Schmidhuber, J. 1997. What's interesting? Technical Report IDSIA-35-97, Istituto Dalle Molle di Studi sull'Intelligenza Artificiale.

Smith, A. J. 2002. Applications of the Self-Organising Map to Reinforcement Learning. Neural Networks (Special Issue), 15: 1107-1124.

Stewart, T. C. 2000. Learning in Artificial Life: Conditioning, Concept Formation, and Sensorimotor Loops. Masters of Philosophy Thesis, University of Sussex.

Streiner, D. 2003. Unicorns Do Exist: A Tutorial on "Proving" the Null Hypothesis. Canadian Journal of Psychiatry, 48 (11): 756-761.

Sutton, R. and A. Barto. 1998. Reinforcement Learning: An Introduction. Massachusetts: MIT Press.

Tani, J. 1998. An interpretation of the 'Self' from the dynamical systems perspective: A constructivist approach, Journal of Consciousness Studies, 5(5/6): 516-542.

Tani, J. and S. Nolfi. 1999. Learning to perceive the world as articulated: an approach for hierarchical learning in sensory-motor systems. Neural Networks 12: 1131-1141.

Touretzky, D.S. and L.M. Saksida. 1997. Operant conditioning in Skinnerbots. Adaptive Behavior 5(3/4): 219-247.

Touzet, C. 1997. Neural Reinforcement Learning for Behaviour Synthesis. Robotics and Autonomous Systems 22: 251-281.

Veogtlin, T. and P.F.M.J. Verschure. 1999. What can robots tell us about brains? A synthetic approach towards the study of learning and problem solving. Reviews in the Neurosciences 10 (3/4): 291-310.