In Proceedings of the Sixth International Conference on Cognitive Modeling, 414-415.
Mahwah, NJ: Lawrence Earlbaum.

World Models, Action Selection, Embodied Concept Formation, and Conditioning

Terrence C Stewart (tcstewar@connect.carleton.ca)
Institute of Cognitive Science, Carleton University
1125 Colonel By Drive, Ottawa, ON, K1S 5B6, Canada

Abstract

There are many computational models whose broad purpose is to allow an agent to learn via experience to perform effectively in a given environment. However, it is uncommon to see these models directly compared to each other, or to empirical data of real creatures adapting to their environments. Here, a comparison methodology is proposed involving various known results in classical and operant conditioning and concept formation. The project involves examining a broad selection of computational models in various environments, and also mixing and matching components from these different models.

Keywords

Hybrid modeling, reinforcement learning, action selection, conditioning, concept formation, world models, sensory pre-processing, model comparisons, supervised learning, unsupervised learning

Introduction

The goal of many computational modelers is to develop agents that can learn to perform well within their environment. To this end, we have seen a proliferation of diverse algorithms and approaches for solving this task. Furthermore, many of these models draw upon other research to define the components of their model. We see back-propagation systems being used within TD(λ) learning (Tesauro, 1995), Genetic Algorithms used to support Action Selection (Farritor & Dubowsky, 2002), Kohonen SOMs combined with Q-Learning (Smith, 2002), and so on.

This variety of approaches is matched with a variety of test problems. Agents learn to control elevators, play board games, navigate grid-worlds, control physical robots, track visual objects, to name just a few. For each situation, results change depending on the sensory representations chosen, the particular implementations of the model's components, and the method whereby the results are judged.

When reading the published results of these models, one is often led to wonder how well a particular model would fare in a completely different domain. Could Tani's Hierarchical RNNs (Tani & Nolfi, 1999) play backgammon successfully? How well would Sarsa or Q-Learning deal with this or that particular robot navigation task? It is rare to find a wide variety of computational models that have been tested on exactly the same task in exactly the same manner (see Gershenson, 2003 for an exception).

This lack makes it difficult to accurately judge the development of the field. When new models (or new variants of old models) are presented, they are often accompanied by a completely new task. Without a comprehensive system of comparison, it is difficult to know the actual strengths and weaknesses of the models. Furthermore, the capabilities of these models are seldom compared with those of real living creatures.

A Unified Approach

Figure 1 shows the framework to which this research restricts itself. A variety of computational models can be seen as particular implementations of this system. The World Model can be implemented by any Supervised Learning scheme. The Sensory and Action States can either be manually set, or modified via any Unsupervised Learning scheme. There are numerous ways of implementing Action Selection, and the various Reinforcement Learning methods can be used to organize the system as a whole.

An advantage of presenting a common framework is that we can start to directly mix and match aspects of these algorithms. Touzet (1997) presents results of using self-organizing maps to automatically reorganize sensor data within a Q-Learning system. This same technique could clearly be applied other cognitive models. Tani's suggestion of using a self-prediction system to help suggest potential actions (Tani & Yamamoto, 2002), or implementing a bias towards novel situations by rewarding incorrect predictions (ibid.), could also be applied to other computational models than his Hierarchical RNNs.

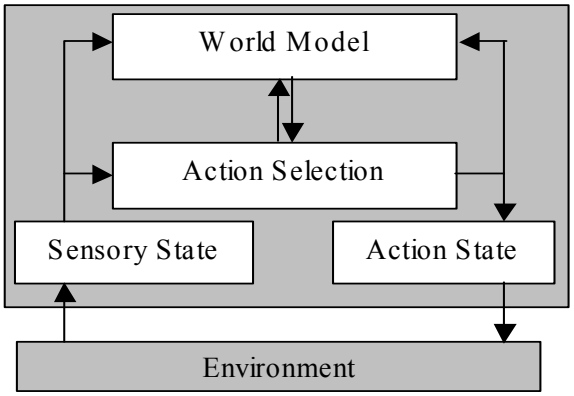

Figure 1: The Generic Agent Model. All models to be investigated fit within this framework.

Most models do not require much coaxing to fit within this framework. They may differ in terms of what exactly is learned by the World Model (does it predict future states or future rewards?). They may have more or less tightly coupled modules (Action Selection systems can make use of neural network world model weights to help find promising actions to perform). But even models as seemingly different as Distributed Adaptive Control or even ACT-R can be seen as falling within this generic framework.

Empirical Data

The existing research literature contains a tremendous variety of tasks and environments that are used to study model performance. From these, we can choose a set of problems to test these algorithms on. Many of the robotic systems cannot be used due to equipment limitations, but any of the simulated worlds (including games and multi-agent systems) are useful domains for comparison.

In addition to these traditional test systems, there is also the potential to compare the performance of these algorithms to the behavior of real living organisms. In (Stewart & Wood, 2001) and (Stewart, 2000), we presented some initial work comparing one model's learning capabilities with the known capabilities of classical and operant conditioning. All reinforcement learning systems could be investigated in these terms. Experiments mirroring those done on living creatures can be performed, measuring rates of acquisition, extinction, specialization, and generalization.

Furthermore, these sorts of comparative studies can also shed light on the ability of the various computational models to form concepts. In the afore-mentioned papers, we identified particular high-level concepts (i.e. regularities of its environment that were not directly represented in the sensory state) that the model was able to identify, and others that it was not able to. This is an embodied approach to concept formation, where an agent is deemed to have recognized an environmental pattern based on its ability to react appropriately to its presence. By combining this with the conditioning research, it is possible to examine those aspects of conditioning which require the modification of concepts (such as generalization and specialization). Existing models of conditioning (such as Kakade & Dayan, 2000) do not discuss such situations.

The Process

The research is divided into three stages. First, the set of supervised systems for developing world models will be compared on a set of test environments. Each model (e.g. ARTMAP, back-prop multi-layer perceptrons, or recurrent neural networks) will be trained using identical input data. The domains will include various grid-worlds, games, and robot navigation environments. The training will be done by giving the model a set of state-action-reward data over time of a particular agent interacting with a world. These various models would then be compared in terms of their ability to predict the future state, reward, or discounted value. Importantly, each model would be trained using exactly the same set of training data. A range of test situations will be used, specifically including ones that would be comparable to real conditioning tasks. For example, Jakobi (1998) discusses a T-maze robotic learning task that is comparable to T-maze tasks given to rodents.

In the second stage, a comparison of full reinforcement learning systems will be performed. Since most of these models make use of supervised learning components to learn the regularities of their environment, we can make use of the results of the previous stage to select promising world model implementations. The environments from the first stage would continue to be used here. The models can be compared to each other based on their resulting total reward. Furthermore, the models' performance can be compared to that of real world creatures in conditioning experiments. For example, if an animal was trained to associate a particular stimulus with a reward, a characteristic acquisition curve would appear in its responses. If this association was extinguished by presenting the stimulus without the reward, a characteristic exponential decay in response should occur. After this, if time passes and then the stimuli is presented again without the reward, the animal will still 'spontaneously recover' the association to a certain degree. This sort of experiment can be performed on any of the reinforcement learning models, allowing us to compare them to the behavior of real living creatures.

In the final stage, various ways of extending this learning model can be investigated. One of these is the use of unsupervised learning to automatically adjust the sensory and motor representations. ARTMAP is an early example of this, and Touzet (1997) gives promising results on using Kohonen SOMs to improve Q-Learning. To perform this comparison, we repeat stage two, with various approaches to automatically pre-processing the sensory data.

References

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

Gershenson, C. (2004). Cognitive paradigms: which one is the best? *Cognitive Systems Research, 5:2*, 135-156

Jakobi, N. (1998). Minimal Simulations for Evolutionary Robotics. PhD thesis, University of Sussex.

Kakade, S. & Dayan, P. (2000). Acquisition in Autoshaping. *Adv. in Neural Information Processing Systems, 12*.

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

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

Stewart, T.C. & Wood, S. (2001). Conditioning and Concept Formation in Embodied Agents. *AAAI Spring Symposium*.

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

Tani, J. & Yamamoto, J. (2002). On the dynamics of robot exploration learning, *Cognitive Systems Research, 3:3*, 459-470.

Tesauro, G. J. (1995). Temporal difference learning and TD-Gammon. *Communications of the ACM, 38*, 58-68.

Touzet, Claude F. (1997). Neural Reinforcement Learning for Behaviour Synthesis. *Robotics and Autonomous Systems, Special issue on Learning Robot: the New Wave*.