In Proceedings of the Sixth International Conference on Cognitive Modeling, 394-395.
Mahwah, NJ: Lawrence Earlbaum.

State Definition in the Tetris Task: Designing a Hybrid Model of Cognition

V. Daniel Veksler (vekslv@rpi.edu)
Wayne D. Gray (grayw@rpi.edu)
Cognitive Science Department
Rensselaer Polytechnic Institute

In defining state/action pairs for reinforcement learning of the Tetris task, we seek to recognize known game states, as well as to learn new ones according to relevant game features. We propose a model of cognition that uses categorization as a mechanism to sort such features into appropriate state-types, and an attention mechanism based on predicted values of each state as a method for deciding which states/features are relevant.

Reinforcement Learning, Categorization, and Spreading Activation

Models of reinforcement learning (RL) have proven to be successful in learning and predicting values of actions given a particular state of the world (Sutton & Barto, 1998). Models of categorization, on the other hand, have been shown to learn and recognize world states with great precision (Medin & Schaffer, 1978). Finally, models of spreading activation, such as ACT-R, have had the most success in matching human memory recall time and accuracy, as well as explaining human priming capabilities (Anderson & Lebiere, 1998).

We have developed a theory for a general model of cognition based on these three mechanisms – CB-Mineral. CB-Mineral (CB) is a Categorization-Based Memory Network with Reinforcement Learning. It is a specialized neural network in which the number of hidden-layer nodes and their interconnections are not predefined, nor are connection weights trained using supervised learning. Rather, memory nodes, connections, and connection weights are incrementally added, consistent with categorization and reinforcement-learning rules.

The State Recognition Problem

Complex dynamic environments, such as Tetris, pose problems for computational modeling. In pairing action sets with system states, the difficulty for production systems, as well as for stand-alone RL models, is in learning new system states (in production systems like ACT-R, the left-hand side, "If X", of any production, "If X, Do Y", is the qualitative equivalent of State in a State/Action pair). It would be both awkward and cognitively implausible for a modeler to hardcode every possible qualitatively unique world-state for a dynamic environment, even one as simple as Tetris.

Categorization models and various connectionist approaches do very well at mapping abstract features from the environment to distinct categories. However, even our hybrid model that uses such methods to establish the state of the world prior to using RL still requires some sort of a heuristic mechanism to decide when categories are created.

CB Implementation

The CB memory network starts its lifecycle with just the perception and the action nodes. Perception nodes are assigned to represent individual contours and locations in the world, as well as pleasure and pain. Action nodes are assigned to various keystrokes and internal commands.

At every tic, the perception module sends signals to the perception nodes, which then activate their subclass nodes. The activation spreads according to connection weights down to the action nodes, which act directly upon the world.

Categorization

Currently, all highly activated objects are considered to be in Working Memory. CB creates a new memory node for every distinct set of objects in WM. This memory node becomes a subclass to each of the objects in WM (Ex: {—} + { | } = {—, | }). If two or more objects in WM share common features, a superclass object is formed for these objects (Ex: {■, location 1, moving down} ∧ {■, location 2, moving down} = {■, moving down}). Each object can have superclasses and subclasses. For example, each Tetris piece (tetronimo) is a subclass to its contours, and is a superclass to its rotation states and falling patterns.

Action Values, State Values, and Connection Strengths

Unlike the traditional RL algorithms that only have a single reward value regardless of whether it is negative or positive, CB records both pain and pleasure values, similar to the way ACT-R records successes and failures.

The value of an action depends on the values of the states that follow that action and the strengths of connections to those states. A given state's value (love/fear), in turn, is based on its connection strength to the pleasure and pain nodes.

The strength of connection between any two memory nodes is determined by the strength of activation of the two nodes at the time when both are in WM at the same time.

Current Work

It has become apparent that not all highly activated memory nodes should be placed in WM. We are currently looking to construct a separate WM buffer that would serve as the blackboard of the mind. The memory elements placed in this buffer would be only the most relevant (according to some heuristic other than activation level) of the highly activated objects.

We refer to Attention as the mechanism responsible for deciding which of the activated memory objects are sent to WM. CB would attend only the most loved and the most feared of the activated memory elements. This would drive the model to reflect upon and learn about only the most relevant task features.

Learning Tetris States

At the start of the game, CB would only "see" a blank board, the contours of the border, and the first falling tetronimo (see Figure 1). Each piece of the contour activates a perception node pre-assigned to detect that particular shape at that particular location, as well as a perception node for that location, and one for that shape.

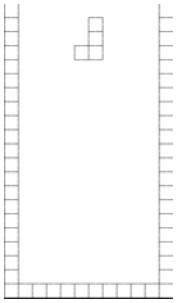

Figure 1. CB Tetris board.

The subclass created out of the co-activated contours makes up what we can semantically translate as that tetronimo (in that rotation state). As the tetronimo moves down one notch, the new combination of active perception nodes is recorded as another object, and so forth. The multiple instances of the tetronimo become the superclasses of an object that may be semantically translated as the "tetronimo falling pattern".

As this pattern is observed multiple times, all the superclass-subclass connections in the memory network are strengthened. With enough connection strength, each "tetronimo" memory object will be activated with just a few distinctive features, and that, in turn, will activate the "tetronimo falling pattern" object. With the creation of the WM buffer, each such object will be pushed into WM for reflection, and all its superclasses – the instances of the falling pattern – will be activated, as well.

Imagination and Planning

The ability to "imagine" future positions of a pattern, as described above, may be the key to planning and decision-making. The model could then keep the bottom of the Tetris board in WM, while mentally going through the various possible states of the current tetronimo in search of a combined state with highly valued responses (as per the rules of RL). The opposite planning pattern may occur as well – the model could also keep the tetronimo in WM, while going through possible states of the board. These two search patterns would be the decision searches for CB in the Tetris task.

Data

Both of the above CB planning patterns were observed in an informal analysis of human eye data during the Tetris task. Collected eye-gaze data suggests that after every newly observed state, subjects quickly search for a complementary object.

After gazing at a new pattern at the bottom of the board, subjects quickly gaze at the top piece to see if it fit into the last observed "hole" in that pattern. Subjects also searched through the bottom of the board for an appropriate "hole" immediately after observing the new tetronimo at the top, with quicker response times when the "hole" matched the tetronimo. We are still looking to do more rigorous analysis of the collected data prior to forming any conclusions.

Summary

In combining the state recognition capabilities of categorization with state prediction of spreading activation and trial-and-error learning of RL, we design the CB-Mineral memory network. The Working Memory buffer and the Attention mechanism are currently being implemented to complete the architecture. Upon completion, CB will be capable of exploring and mastering Tetris as a sample complex dynamic environment. A preliminary look at human eye data suggests a qualitative match to the predicted cognitive flow of the model, leaving the doors open for future research.

References

Anderson, J., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, Massachusetts: The MIT Press.