Feature-based learning increases the generalizability of state predictions

Euan Prentis
Department of Psychology
University of Chicago
eprentis@uchicago.edu

Akram Bakkour
Department of Psychology
University of Chicago
bakkour@uchicago.edu

Abstract

Decisions have consequences that gradually unfold over time. To make effective decisions, it is therefore necessary to learn not only which states of the world are useful to be in (*value-based learning*) but also whether these states will be visited in the future (*state-predictive learning*). However, in real-world contexts, states are complex and vary along numerous feature dimensions. This reduces the likelihood that a given combination of features will reoccur, in turn limiting the extent to which past learning can be applied to relevant future experiences. This problem is known as the *curse of dimensionality*. Feature-based learning has been shown to mitigate the curse of dimensionality in the domain of pure value-based learning [1]; theoretically, feature-based learning should improve learning speed, generalizability, and compositionality. The present work addresses whether these advantages extend to the realm of predictive learning. We implement state- and feature-based successor representation models, and simulate their behavior on a novel sequential learning task in which sequences can be learned at either the state or feature level. We found that feature-based learning improves the speed, generalizability, and compositionality of predictive learning. Varying the amount of training each model received, we additionally observed that these advantages were most pronounced with less training. These results support the notion that feature-based learning (1) facilitates quick generalization in novel sequential learning problems, and (2) has the potential to mitigate the curse of dimensionality in real-world contexts. Continuing work will adapt the described task to probe whether humans use feature-based learning to make predictive inferences.

Keywords: Feature-based Learning, Successor Representation, Human Cognition, Compositionality

1 Introduction

Decisions have consequences that gradually unfold over time. For example, after choosing to drive to work instead of taking the subway, you may run into rush-hour traffic, fail to find parking near your workplace, arrive to work late, and ultimately get scolded by your boss for your tardiness. To make effective decisions, it is therefore necessary to learn not only which states of the world are useful to be in (e.g., at work on time; *value-based learning*) but also whether these states will be visited in the future (*state-predictive learning*).

States in real-world contexts are complex and high-dimensional. For example, as you consider which mode of transport to take to work, there will be differences in the weather, your mood, how long you slept, and so forth. This renders state-predictive learning non-trivial to perform. Due to variance along many numerous dimensions, it is unlikely that the exact same combination of features will reoccur. Learning predictions at the multidimensional state level therefore limits the extent to which learning can be applied to relevant future experiences. This problem is known as the *curse of dimensionality*.

In the domain of pure value-based learning, the curse of dimensionality can be mitigated by learning on the decomposed features of states (feature-based learning) rather than the states themselves (state-based learning, [1, 2, 3]). Theoretically, this improves learning along three dimensions: (1) learning speed, (2) generalizability, and (3) compositionality. (1) *Learning speed* is faster because tasks generally have fewer features than states, meaning that features are encountered more frequently. This allows for more opportunities to learn over the same number of training iterations. (2) *Generalizability* is greater because a single feature can be present across multiple states. Therefore, anything learned about a given feature in one state can be applied to any novel state partially composed of the familiar feature. (3) *Compositionality* is greater, since decomposed features encountered in different contexts may be re-combined into novel configurations with some inferred value. An agent can harness the power of compositionality to generate creative, goal-oriented action plans.

The present work addresses whether the benefits of feature-based learning extend to the realm of predictive learning. Using successor representation modeling, we simulate the behavior of state- and feature-predictive learners, and compare the speed, generalizability, and compositionality of their learning.

2 Successor Representation Modeling

State-predictive learning can be modeled using the successor representation (SR, [4]). SR learns a compressed version tasks' multi-step transition structures. This facilitates inference about distant outcomes without performing computationally intensive tree searches through the full state-space (e.g., as with model-based reinforcement learning [5]). Since searching through the large decision trees of real-world contexts is likely intractable, the computational efficiency of SR makes it a good candidate for human state-predictive learning. Work identifying signatures of SR in human behavioral and fMRI data supports this theory [6, 7, 8, 9].

SR learns estimated values and state predictions independently through temporal difference learning [10]. After a reward r is observed, the current state's estimated value V_s is updated according to the reward prediction error, plus the discounted estimated value of the next state $V_{s_{new}}$:

$$V_s = V_s + \alpha(r + \gamma V_{s_{new}} - V_s) \tag{1}$$

where free parameters α and γ respectively control the learning and discount rates. State predictions are represented in the successor matrix $M \in \mathbb{R}^{S \times S}$, where rows and columns correspond to current and new states, respectively. After a transition is observed, a count $e_{s_{new}}$ is added to the new state's position in the row vector. Since this update incorporates the discounted predictions of the new state, SR gradually learns to predict distant visitations. Formally:

$$M_s = M_s + \alpha(e_{s_{new}} + \gamma M_{s_{new}} - M_s) \tag{2}$$

When evaluating given state s, both the estimated values and visitation expectancies are incorporated:

$$O_s = M_s V_s \tag{3}$$

State values O_s are then turned into choice probabilities using a softmax with inverse temperature β:

$$p(a|s, s') = \frac{e^{\frac{O_s}{\beta}}}{e^{\frac{O_s}{\beta}} + e^{\frac{O_{s\prime}}{\beta}}} \tag{4}$$

2.1 State-based SR

State-based SR learns and evaluates actions as described. It additionally caches visited states in memory as it interacts with the environment. To generalize to a novel state, the model identifies the most similar representation in memory, and retrieves the associated V_s and M_s. State-based SR also uses representational similarity to guide composition. When tasked with combining decomposed features into some rewarding state, the model considers all possible combinations of the provided features, and deterministically builds the combination which is most similar to high-value states in memory.

2.2 Feature-based SR

Feature-based SR is achieved by simply tweaking the model to learn V and M separately for each feature category f. In this case, final values F_{fi} are produced per feature instance i.

$$F_{fi} = M_{fi}V_{fi} \tag{5}$$

The state's final value O_s is the mean of these feature values. To generalize to novel states composed of partially familiar features, feature-based SR directly calculates O_s from the familiar feature values F_{fi}. When tasked with combining decomposed features into some rewarding state, the model also directly retrieves associated F_{fi}'s and deterministically builds the state that maximizes these values.

3 Sequential Learning Task

We implemented a sequential learning task (Figure 1), in which agents' goal was to maximize point earnings. The task consisted of three phases: training, test, and composition.

During training, choices were made between pairs of items that represented different states (Figure 1A). Each item was composed of three feature instances, sampled from five feature categories (ABCDE). Only three of the five categories would be seen in each training half (1st half: ABC; 2nd half: ADE; Figure 1C). After a choice was made, a single-step sequence was displayed followed by a reward (r = [-6, 6]). The successor item's

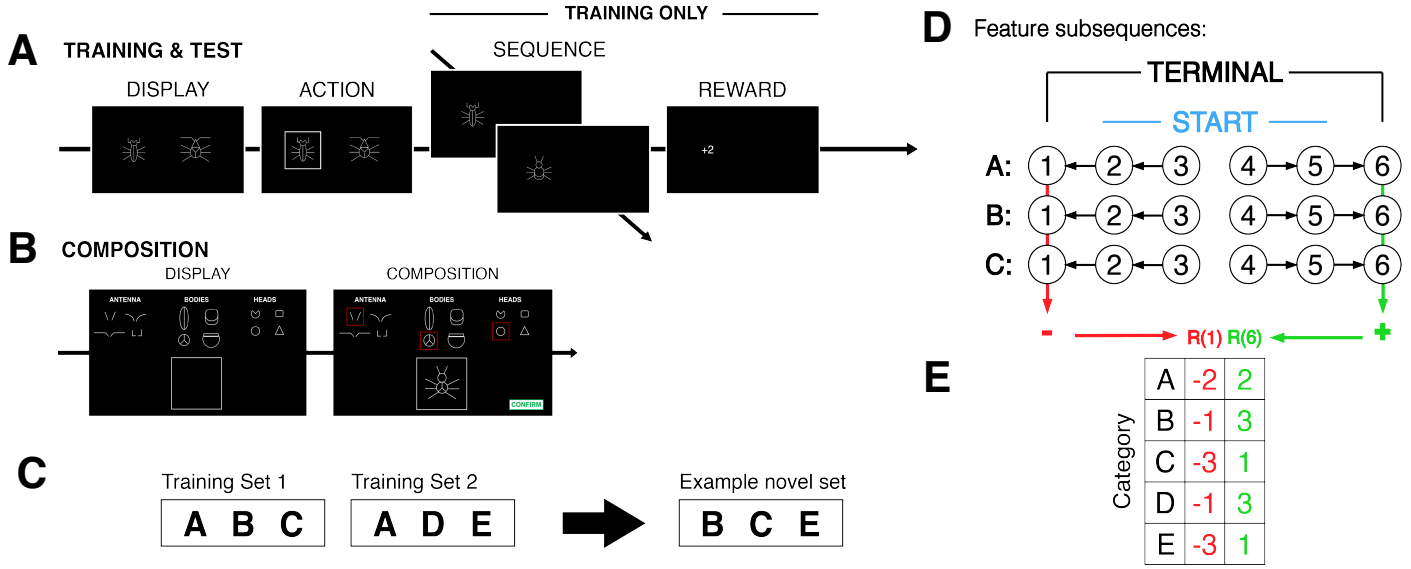

Figure 1: Task Design. A. Training and test trials. After an action was made during training, a one-step sequence and reward were observed (these were not seen at test). Example stimuli here are from the human subjects task. **B.** Composition trial. Agents selected decomposed features from different categories on each trial. **C.** Feature category sets. Whereas training items were directly sampled from a fixed set in each training half (i.e., ABC or ADE), novel items were composed of a mixture of features across halves (e.g., ACE, BDE, BCE). **D.** Sub-sequences associated with each feature category. Start items were composed of feature instances 2, 3, 4, or 5. Successor items were partially or fully composed of terminal feature instances (1, 6), that are associated with a reward value. **E.** Reward values of terminal feature instances. The reward values for terminal feature instances associated with a successor item would be summed to get the item reward value.

identity and reward value were determined by sub-sequences associated with each feature category (Figure 1DE). Thus, to maximize cumulative reward, agents had to predict which choices would lead to rewarding successor items.

Test was similar to training, but agents neither observed the one-step sequences nor the reward outcomes, preventing further learning. On each test trial, choices were between either training or novel items (Figure 1C). We hypothesized that feature-based learning is more generalizable, and thus expected the feature-based model to perform better on novel item trials. Trials also differed on whether choices were between terminal or start items (Figure 1D). In contrast to terminal items, start items were never worth points, and only indirectly led to reward outcomes through terminal items. Therefore, to make accurate inferences on start item trials, agents needed to apply both value and state-predictive learning.

Finally, agents completed a composition phase. On each trial, they were presented with decomposed feature instances, and had to recompose them into a reward-predictive item (Figure 1B). These trials also differed on whether training or novel items, and start or terminal items could be composed. Since the feature-based model learns directly on features, we predicted that it would be more effective at constructing rewarding items than the state-based model.

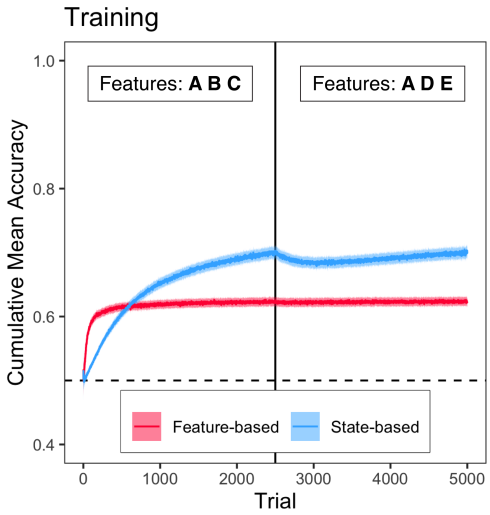

Figure 2: Training Curves. Curves are quantified by cumulative mean accuracy, and averaged over 1000 simulations of 5000 trials per model. Errors at 95% confidence intervals. The first 2500 trials involved features from categories ABC, and the remaining trials involved features from categories ADE.

4 Results

We simulated the feature- and state-based learning models on the sequential learning task, and compared the speed, generalizability, and compositionality of their learning.

First, we probed whether feature-based learning is faster by comparing the models' learning trajectories over 5000 training trials. We calculated the cumulative mean accuracy (CMA) of each agent's choices, where an accurate choice is defined as one that will lead to a more rewarding successor item. In each training half, since feature-based learners encountered each of the 18 feature instances more frequently than the state-based learners encountered each of the 128 unique items, we hypothesized that feature-based learners would achieve higher accuracy earlier in training. The results reflect this hypothesis (Figure 2). Whereas state-based learners surpassed a mean CMA of 0.6 on trial 465, feature-based learners surpassed a mean CMA of 0.6 on trial 150. However, on the final trial, state-based learners achieved a higher mean CMA (M = 0.70) than the feature-based learners (M = 0.62). These results indicate that while feature-based learning is faster, state-based learning is more accurate in the long run. There was also a more substantial drop in state-based learners' accuracy entering the second training half (where feature categories B and C were replaced with D and E). This indicates they were poorer at applying past learning to

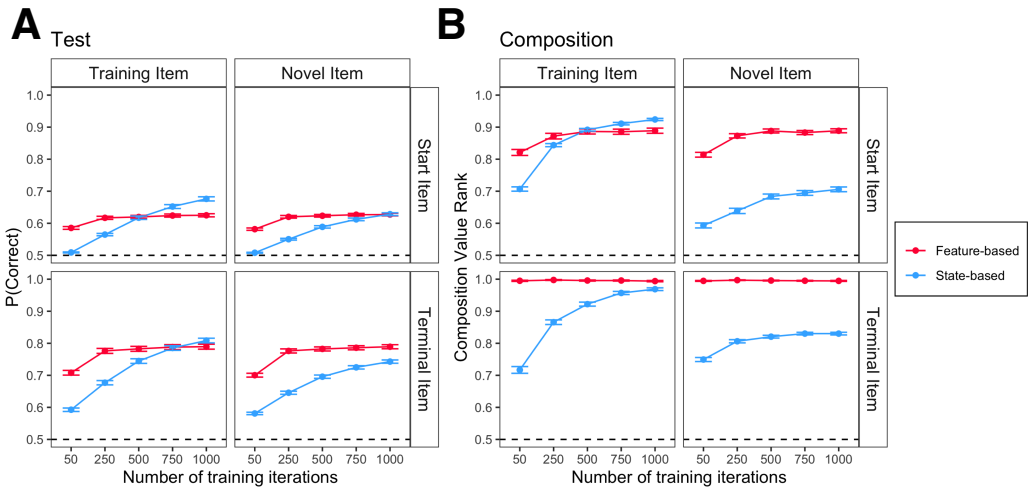

Figure 3: Test and Composition Performance. Errors are 95% confidence intervals. **A.** Proportion of correct choices made during test, by training amount and item type. **B.** Composition-value rank, by training amount and item type.

the new context and learning about the novel items.

Next, we assessed the models' abilities to generalize during the test phase with different amounts of training (50, 250, 500, 750, or 1000 trials). As predicted, we found that the feature-based model was better at generalizing, particularly with fewer training iterations (Figure 3A). With just 50 training iterations, the feature-based learners were more accurate on all trial types. However, the magnitude of this difference reduced with more learning, supporting the notion that while feature-based learning is faster, state-based learning performs better in the long-term. Notably, whereas the state-based learners' training accuracy came to surpass that of feature-based learners, their novel item accuracy did not. This suggests feature-based learning particularly facilitated generalization.

Lastly, we compared the models' performance at composing reward-predictive items during the composition phase. High performance is operationalized as the reward-value rank of the composed item, relative to all other possible combinations of features presented on a given trial. We predicted that the feature-based model would be able to leverage its feature-level learning to compose rewarding items more precisely, and thus achieve higher overall performance. The results support this hypothesis (Figure 3B). Notably, feature-based learners achieved higher composition accuracy for novel items across all numbers of training iterations.

5 Discussion

Simulating feature- and state-based predictive learning, we found that learning on the features of states rather than the states themselves bolsters learning speed, generalizability, and compositionality. These advantages were particularly pronounced with less training. Our results have two important implications.

Firstly, in novel tasks, agents would benefit from performing feature-based learning to achieve some degree of accuracy quickly. However, with extended experience, they may benefit from switching to more state-based mechanisms. Farashahi, Rowe, and colleagues [1] demonstrated similar advantages in the domain of pure value-based learning, finding that human subjects gradually switched from feature- to state-based learning in a low-dimensional environment. State abstraction also increases the generalizability of predictive learning [11]. However, since abstractions must be learned over multiple experiences, feature-based learning may be advantageous prior to the formation of stable abstractions.

Secondly, feature-based learning may help agents learn structural contingencies in real-world environments, where the curse of dimensionality renders state-based learning ineffective. For this reason, human beings may rely on feature-based learning to make predictive inferences. Continuing work will adapt the sequential learning task described here to test this hypothesis.

References

1. Farashahi, S., Rowe, K., Aslami, Z., Lee, D., & Soltani, A. (2017). Feature-based learning improves adaptability without compromising precision. *Nature Communications*, *8*(1), 1768.
2. Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron*, *93*(2), 451–463.
3. Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement Learning in Multidimensional Environments Relies on Attention Mechanisms. *Journal of Neuroscience*, *35*(21), 8145–8157.
4. Dayan, P. (1993). *Improving Generalisation for Temporal Difference Learning: The Successor Representation*. 14.
5. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215.
6. Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behaviour*, *1*(9), 680–692.
7. Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2021). *Neural evidence for the successor representation in choice evaluation* [Preprint]. Neuroscience.
8. Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, *20*(11), 1643–1653.
9. Gershman, S. J. (2018). The Successor Representation: Its Computational Logic and Neural Substrates. *The Journal of Neuroscience*, *38*(33), 7193–7200.
10. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (Second edition). The MIT Press.
11. Lehnert, L., Littman, M. L., & Frank, M. J. (2020). Reward-predictive representations generalize across tasks in reinforcement learning. *PLOS Computational Biology*, *16*(10), e1008317.