Optimal Learning in Temporally Structured Environments

Niloufar Razmi (niloufar_razmi@brown.edu)
Department of Neuroscience, Brown University
Providence, Rhode Island, USA

Matthew Nassar (matthew_nassar@brown.edu)
Department of Neuroscience, Brown University
Providence, Rhode Island, USA

Abstract

Biological agents live in a dynamic world and can exploit structure in their environment to improve the efficiency of learning. Previous work has yielded normative learning algorithms that prescribe learning strategies for specific environmental structures, but leave open the question of how humans and animals might infer the structure of their current environment. In this project, we propose an optimal theoretical model of learning the structure of varying environments. Specifically, we define learning the structure of change as putting a prior on the transition matrix of a hidden Markov model and using observations to update that prior with Bayes rule. With minimum assumptions imposed on the generative model of the environment statistics, we test our model in four different environments and find signatures of context-appropriate behavior previously observed in humans. Our work proposes the first unifying model of adaptive learning through experience in complex temporally structured environments.

Keywords: Structure learning; State representation; Bayesian inference; Hierarchical Dirichlet process.

Past research in human learning and decision making, ranging from systems neuroscience to neuroeconomics, has suggested that people adaptively adjust learning in response to a changing environment. Bayesian models in particular, have been successful to provide insight into learning adjustments in a given environment. Nevertheless, we still lack an understanding of how people, or even algorithms, might come to understand the underlying temporal structure of different environments without explicit instruction. In this extended abstract, we first explain the general framework of the tasks we studied, then explain our model, the Hierarchical Dirichlet structure learner (HDSL) and lastly present the results of the modeling.

We consider an online prediction task in which a stream of data points x_t arrive at discrete time points t, and the task of the learner is to predict $p(X_t|X_1,...,X_{t-1})$. We generate tasks by using a hidden Markov model. Specifically, we simulate four canonical statistical structures by using four types of transition probability matrix of the HMM where the rows are probability of transitions from state Z_i to state Z_j with Z denoting a categorical "state" (Figure 2A).

Our modeling approach consists of choosing an appropriate prior for prediction of outcome of a new trial and using Bayes rule to update that prior with each new observation. We use a Hierarchical Dirichlet process to generate the transition probability matrix of our HMM, enabling our model to scale to, in principle, an infinite number of states.

We generate the prior from Hierarchical Dirichlet Process in three steps:

1) first we draw a vector β from a Dirichlet Process (DP) where $\sum_i \beta_i = 1$. The vector β is called the global transition probabilities and denotes the probability of each state in the hidden Markov model.

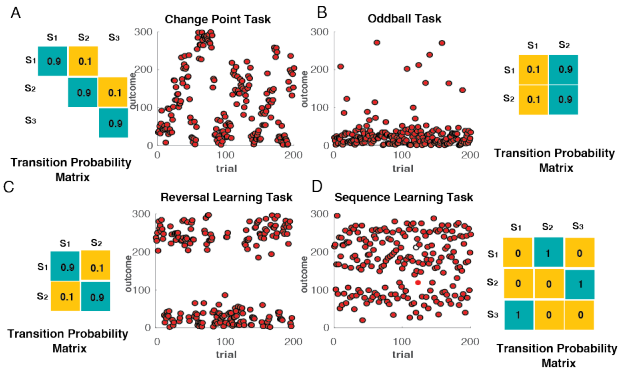

Figure 1: We studied four canonical temporal structures often examined in human empirical studies: A) change points, B) oddballs, C) reversals and D) sequences. Each structure is defined by qualitative features of the state transition matrix (left) which gives rise to discontinuities in observations over time (right). Though each task incorporates discontinuous temporal dynamics, they each elicit unique learning strategies in human participants and normative models.

We generate these probabilities via a stick-breaking construction where we recursively draw a random variable β'_k from $Beta(1,\gamma)$, Higher values of gamma results in smaller broken stick portions and higher probability of new states in the future, thus this hyper parameter would determine model complexity.

2) When a new state is created, we generate the corresponding row in the transition probability matrix T by using the original β vector as the base of a second DP :

$$T_j \sim DP(\alpha,\beta) \tag{1}$$

Where α is the concentration parameter and β is the base vector of the Dirichlet process. To generate rows of the transition probability matrix according to we use a Chinese restaurant process where the probability of transitioning from each state i to previously visited state j is s:

$$P(Z_t = j|Z_{t-1} = i) = \frac{\alpha\beta_j + n_{ij}}{\alpha\beta + N} \tag{2}$$

where $n_i j$ is the number of previous transitions from state i to j and N is the total number of transitions (i.e. trials) observed so far. If j is a new state, the probability of transitioning to it is given by:

$$P(Z_t = j|Z_{t-1} = i) = \frac{\alpha\beta_j}{\alpha\beta + N} \tag{3}$$

3) Lastly, to control the dynamics of self transitions, we take a weighted average of the transition matrix obtained from steps 1 and 2 and the identity matrix of the same size (i.e. a matrix with ones on the diagonal and zeros everywhere else), where the weight of the second matrix is η. The third hyper parameter of the HDP, η, is seen as a persistence factor that favors higher probabilities of self-transitions (the diagonal of

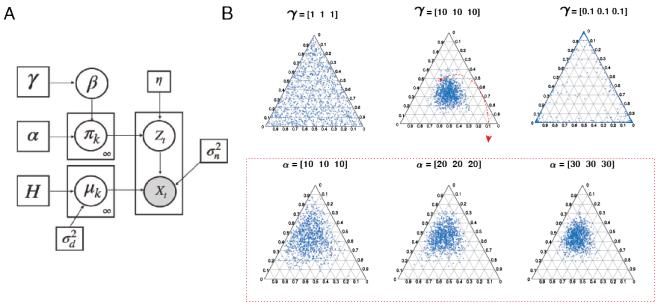

Figure 2: Graphical model of the hierarchical Dirichlet process prior described in the text. B) Gamma and Alpha act as inverse variance-like hyper parameters. Example draws from a 3 dimensional DP: 1) Top row: In the first level DP, gamma controls the overall probability of transitioning to each discrete state 2) Bottom row: alpha controls the similarities of draws from the second DP to one single draw from the distribution in top row, middle panel. Thus, alpha controls how similar each column of the transition matrix is to all of the others.

the transition probability matrix). The generative process described here is depicted in the graphical model in Figure 2 B.

For making inference on the generative model explained in previous section, we used an online particle filtering approach for approximating the posterior at each time step. Specifically, we start with t=1 and n equally weighted particles uniformly distributed on an equally spaced 5-dimensional grid of the free parameters of the model. Each particle is identified by a unique combination of free parameters = $\{\sigma_n^2, \sigma_d^2, \alpha, \gamma, \eta\}$. Each particle is then copied p times. These copies share the same hyper parameters and, as a group, are used to approximate a probability distribution over the sequence of states. The inference model observed task data one trial at a time, using standard particle filtering equations and our inverted generative model for learning, and model predictions were extracted as the maximum likelihood estimates of the underlying mean on the upcoming trial.

We first show that the model is able to learn all of the four prototypical tasks compared to other existing models in the literature (Figure 3). But good performance in the tasks doesn't necessarily mean the model is showing adaptive behavior. For example, in many tasks, heuristics can achieve good performance without taking any of the underlying environmental structure into account. Thus, we analyzed the qualitative signatures of adaptive behavior for each of the task types and showed that the model's reaction to change is consistent with learning the temporal structure of each task (Figure 4).

How biological agents are able to learn the underlying structure of a changing environment is still unknown. We are first to provide a generalized Bayesian optimal model that is able to learn temporal structure of different kinds of environment statistics in a unifying way.

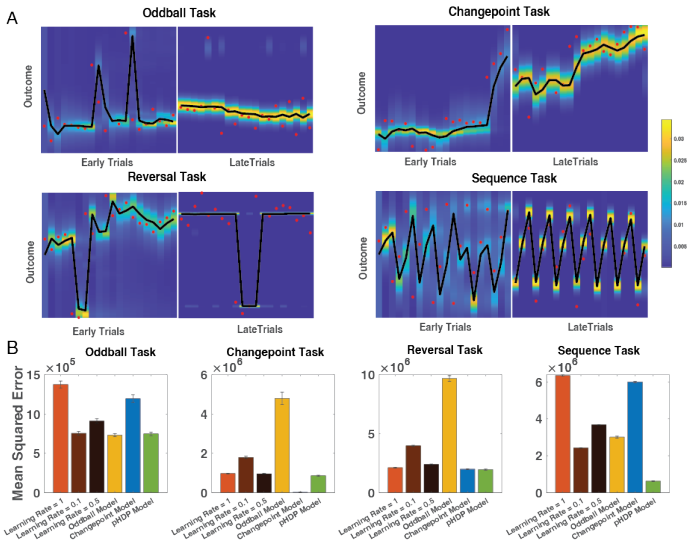

Figure 3: The results of the online Bayesian inference model with a HDP prior in the four different task types studied here. A) The outcome locations (0-300) on each trial is depicted in red circles with the full predictive distribution (the heat map) and the model prediction based on the maximum value of the predictive distribution (solid black line) for oddball, change point, reversal and sequence learning tasks, comparing early trials (first 20 trials) and late trials (last 20 trials) B) Comparison of the performance of our model with: three models with fixed learning rates and Bayesian models optimal for the changepoint and oddball tasks. Even though some models might perform better on a single task, our Bayesian model with the HDP prior does well across all of the tasks.

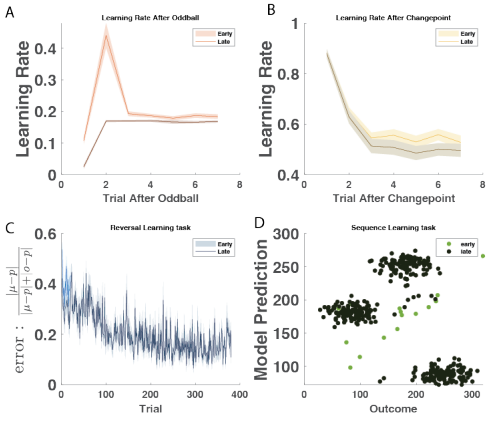

Figure 4: Our Bayesian model with the HDP prior shows qualitatively similar behavior previously observed in human participants. Namely, a high learning rate after change points (B) but not after oddballs(A). Gradually decreasing error (defined as the difference between prediction and outcome mean) in the reversal learning task (C) and learning to predict the next trial mean instead of current trial outcome in the sequence learning task (D).

References

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nat. Neurosci.*, *10*(9), 1214–1221.

Heald, J. B., Lengyel, M., & Wolpert, D. M. (2021). Contextual inference underlies the learning of sensorimotor repertoires. *Nature*, *600*(7889), 489–493.

Razmi, N., & Nassar, M. R. (2022). Adaptive learning through temporal dynamics of state representation. *J. Neurosci.*, *42*(12), 2524–2538.

Tavoni, G., Doi, T., Pizzica, C., Balasubramanian, V., & Gold, J. I. (2022). Human inference reflects a normative balance of complexity and accuracy. *Nat. Hum. Behav.*, *6*(8), 1153–1168.

Yu, L. Q., Wilson, R. C., & Nassar, M. R. (2021). Adaptive learning is structure learning in time. *Neuroscience Biobehavioral Reviews*, *128*, 270-281.