Inverse stochastic learning

Momchil S. Tomov (mtomov@g.harvard.edu)
Department of Psychology and Center for Brain Science, Harvard University, Cambridge, MA
Motional AD LLC, Boston, MA

Abstract

Rational process models posit that the brain learns the hidden structure of the world by approximating Bayesian inference using Monte Carlo sampling. The stochasticity of such inference algorithms makes it challenging to study the neural basis of the learning process, since it is difficult for the experimenter to know what the subject has inferred at any point in time. Here we tackle this inverse learning problem within the framework of inverse rational control using a simple particle filtering scheme. We evaluate our method on synthetic data and show that it uncovers the hidden states inferred by a subject on a trial-by-trial basis more accurately than a generative approach that only simulates the learning process. We then discuss how this method could be applied to a wide range of topics in cognitive computational neuroscience.

Keywords: Bayesian inference, Monte Carlo, structure learning, POMDP, MCMC, particle filtering, bounded rationality, rational process models, inverse rational control

Introduction

A hallmark of intelligence is the ability to rapidly adapt to new tasks based on past experience with similar tasks. This "learning-to-learn" is thought to be supported by the brain's ability to construct a mental model of the hidden structure of the world that abstracts away particular experiences (Braun, Mehring, & Wolpert, 2010). Such *structure learning* encompasses a broad range of phenomena, including latent learning (Tolman, 1948), formation of learning sets (Harlow, 1949), causal inference (Bramley, Dayan, Griffiths, & Lagnado, 2017), and theory learning (Ullman, Goodman, & Tenenbaum, 2012). Studying the neural basis of structure learning often involves linking the hidden structure inferred by a subject during an experiment with simultaneously recorded neural activity (Tomov, Dorfman, & Gershman, 2018). Yet from the point of view of the experimenter, the structure inferred by the subject is itself a latent variable that must be inferred.

Broadly, there are two main approaches to inferring latent variables represented by the brain: bottom-up (or data-driven) approaches and top-down (or theory-driven) approaches (Linderman & Gershman, 2017). Data-driven approaches, such as state space models, start with weak assumptions about how the latent variable evolves over time and let the data fully determine it (Wiltschko et al., 2015). In contrast, theory-driven approaches, such as ideal observer models, start with strong assumptions about the learning process, usually by formalizing it as Bayesian inference (Griffiths & Tenenbaum, 2005). Data-driven approaches tend to be underconstrained, potentially overfitting to noise and resulting in inferences that are difficult to interpret. Theory-driven approaches, on the other hand, tend to be too constrained, often treating deviations from optimal behavior as statistical noise. Further, since perfect inference is often intractable, theory-driven approaches often require approximations, which potentially render the results even less biologically plausible.

Contrary to the last point, a number of researchers have proposed that the stochastic firing of neurons in the brain might, in fact, implement such sampling-based approximations of Bayesian inference (Gigerenzer, 1993; Buesing, Bill, Nessler, & Maass, 2011; Gershman, Horvitz, & Tenenbaum, 2015). In this view, structure learning can be cast as a kind of Monte Carlo sampling which stochastically explores the space of possible structures and asymptotically converges the hidden structure that is most consistent with the observations (Sanborn, Griffiths, & Navarro, 2010). Despite its cognitive plausibility, this approach further complicates the analysis of neural data, as the structure inferred by the model could deviate substantially from the structure inferred by the brain. This leaves open the question of how to recover the time course of subjective beliefs about the structure of the world from data.

Methods

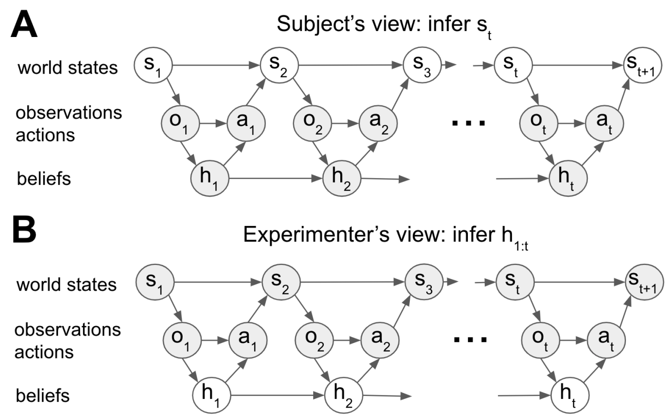

Figure 1: The learning problem from subject's point of view (A) and the inverse learning problem from experimenter's point of view (B). Adapted from Wu et al. (2018). s, world state/structure. o, observation. a, action. h, subject's belief about world state/structure. Empty circles denote latent variables, solid circles denote observable variables.

To address this question, we extend the framework of inverse rational control (Wu et al., 2018) to accommodate stochastic belief updating. In particular, we make the following assumptions (Figure 1):

- The world has some hidden structure – a kind of abstract state s – that is unobservable to the subject.
- The subject can infer this hidden state based on their observations o and actions a using Bayes' rule: $P(s|o,a) \propto P(o|s,a)P(s)$.
- The subject approximates this posterior using a single hypothesis h which is updated stochastically, for example using Markov chain Monte Carlo (MCMC).
- The experimenter's goal is to infer the sequence of hypotheses $h_{1:T}$ inferred by the subject, given the subject's observations $o_{1:T}$ and actions $a_{1:T}$.

In short, the subject's learning process is modeled as sequential Monte Carlo sampling (Figure 1A). This frames the experimenter's inverse learning problem as an input-output hidden Markov model (HMM, Sahani (2014); Figure 1B).

The dynamics of the HMM are governed by:

$$h_t \sim f(h_t|h_{t-1}, o_t) \tag{1}$$

$$a_t \sim \pi(a_t|h_t, o_t), \tag{2}$$

where the transition distribution f corresponds to the subject's learning process and the emission distribution π is the subject's policy, parametrized by θ (omitted). This way of modeling the subject's beliefs strikes a balance between top-down and bottom-up approaches: it is theoretically-constrained (via Eq. 1) and yet it takes the data into account (via Eq. 2).

```
def filtering(o, a, N, sample_f, pi):
  T = len(a) # number of time points
  particles = [Particle(T) for _ in range(N)]
  w = [0] * N # importance weights
  for t in range(T): # for each timestep
    for i in range(N): # for each particle
      # initialize hypothesis from prior (at t=0)
          or update hypothesis (Eq. 1)
      particles[i].h[t] = init_hypothesis()
        if t == 0
        else sample_f(particles[i].h[t-1], o[t])
      # reweigh hypothesis (Eq. 2)
      w[i] = pi(a[t], particles[i].h[t], o[t])
    particles = resample(particles, w)
  return particles
```

Algorithm 1: Inverse stochastic learning.

We solve the HMM using sequential importance resampling, a kind of particle filtering (Doucet & Johansen, 2009) that approximates the joint posterior over all timesteps T, $P(h_{1:T}|o_{1:T}, a_{1:T})$ (Algorithm 1). We initialize N hypotheses (particles) from a prior. Then, at each timestep, we update each particle following Eq. 1 by sampling according to the subject's learning process f. We then reweigh the particles following Eq. 2 according to the subject's policy π, resample with replacement, and move to the next timestep. We fit the parameters θ using expectation-maximization (not shown).

Intuitively, this scheme amounts to generatively simulating multiple versions (particles) of the subject's stochastic learning process in parallel and pruning particles that are inconsistent with the subject's actual behavior. Note that this only requires the ability to sample from f and to compute π. Despite its simplicity, this approach converges to the true posterior $P(h_{1:T}|o_{1:T}, a_{1:T})$ as $N \to \infty$.

Results

We applied this approach to a simplified simulated reversal learning task (Aguillon-Rodriguez et al., 2021). Simulations allowed us to evaluate our approach against the subject's actual beliefs, which are generally inaccessible to the experimenter. A simulated subject performed a two-alternative forced choice task where, on each trial, either the left or the right choice was rewarded (Figure 2A). The design was blocked, such that in each block, one side

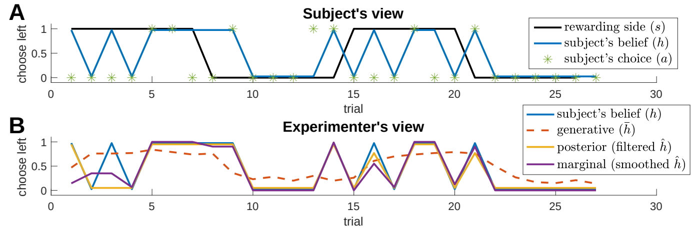

Figure 2: Inverse stochastic learning applied to simulated reversal learning. (A) Subject's point of view: rewarded side $s_{1:T}$, unobservable to subject (black); subject's belief $h_{1:T}$ about rewarded side (blue); subject's choices $a_{1:T}$ (green). (B) Experimenter's point of view: subject's belief $h_{1:T}$, unobservable to experimenter (blue; same as in A); average belief $\bar{h}_{1:T}$ over generative simulations, ignoring the subject's behavior (red); inferred belief $\hat{h}_{1:T}$ based on particle filtering (yellow) and smoothing (purple).

was rewarded. The rewarded side was the hidden state, i.e. $s_t = \mathbb{I}[\texttt{left side is rewarded on trial t}]$. We assumed the subject employs a simple stochastic win-stay lose-shift strategy f that maintains the same belief h following rewarded trials and flips with some probability following unrewarded trials. We also assumed a stochastic policy π that chooses the favored side with some high probability. While technically not structure learning, this example serves to illustrate our method and also highlight its general applicability.

Simulating the subject's learning process generatively (Eq. 1), without taking subject behavior into account (Eq. 2), correlates with the ground truth state s ($r = 0.97 \pm 0.01$) and the subject's belief h ($r = 0.53 \pm 0.02$; Figure 2B). However, the correlation is significantly stronger when using particle filtering ($r = 0.66 \pm 0.02, T(99) = 6.70, p < 10^{-8}$) or smoothing ($r = 0.73 \pm 0.02, T(99) = 11.57, p < 10^{-19}$) that additionally take behavior into account (Eq. 2). Smoothing performs better than filtering alone ($T(99) = 4.03, p = 0.0001$).

Discussion

Our method can be used to recover the trajectory of structural beliefs based on subject behavior, particularly in complex domains in which optimal inference is intractable. By mapping such beliefs to brain activity, researchers can evaluate different hypotheses about how the brain learns the structure of the world. This could be applied to reverse-engineering the neural mechanisms supporting different kinds of structure learning, including causal structure learning (Tomov et al., 2018), social structure learning (Lau, Gershman, & Cikara, 2020), context-dependent learning (Geerts, Gershman, Burgess, & Stachenfeld, 2023), hippocampal remapping (Sanders, Wilson, & Gershman, 2020), and theory learning (Tomov, Tsividis, Pouncy, Tenenbaum, & Gershman, 2023). In addition, our method can be applied to other instances of sampling-based learning, such as concept learning (Goodman, Tenenbaum, Feldman, & Griffiths, 2008), change point detection (Brown & Steyvers, 2009), and conditioning (Daw & Courville, 2008).

Acknowledgments

This research was supported by the Toyota Corporation, the Center for Brains, Minds and Machines (CBMM), funded by NSF STC award CCF1231216, and the Multi-University Research Initiative Grant (ONR/DoD N00014-17-1-2961). We are grateful to Zhengwei Wu, Samuel Gershman, Chris Bates, and Lucy Lai for the fruitful discussions.

Data availability

Code for the simulations can be found at `https://github.com/tomov/ISL-demo`

References

Aguillon-Rodriguez, V., Angelaki, D., Bayer, H., Bonacchi, N., Carandini, M., Cazettes, F., ... Zador, A. M. (2021). Standardized and reproducible measurement of decision-making in mice. *eLife*, *10*, e63711.

Bramley, N. R., Dayan, P., Griffiths, T. L., & Lagnado, D. A. (2017). Formalizing neurath's ship: Approximate algorithms for online causal learning. *Psychological review*, *124*(3), 301.

Braun, D. A., Mehring, C., & Wolpert, D. M. (2010). Structure learning in action. *Behavioural brain research*, *206*(2), 157–165.

Brown, S. D., & Steyvers, M. (2009). Detecting and predicting changes. *Cognitive psychology*, *58*(1), 49–67.

Buesing, L., Bill, J., Nessler, B., & Maass, W. (2011). Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons. *PLoS computational biology*, *7*(11), e1002211.

Daw, N., & Courville, A. (2008). The pigeon as particle filter. *Advances in neural information processing systems*, *20*, 369–376.

Doucet, A., & Johansen, A. M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, *12*(656-704), 3.

Geerts, J. P., Gershman, S. J., Burgess, N., & Stachenfeld, K. L. (2023). A probabilistic successor representation for context-dependent learning. *Psychological Review*.

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278.

Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In *Rationality: Psychological and philosophical perspectives* (pp. 284–313). Routledge.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive science*, *32*(1), 108–154.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive psychology*, *51*(4), 334–384.

Harlow, H. F. (1949). The formation of learning sets. *Psychological Review*, *56*(1), 51.

Lau, T., Gershman, S. J., & Cikara, M. (2020). Social structure learning in human anterior insula. *Elife*, *9*, e53162.

Linderman, S. W., & Gershman, S. J. (2017). Using computational theory to constrain statistical models of neural data. *Current opinion in neurobiology*, *46*, 14–24.

Sahani, M. (2014). Probabilistic & unsupervised learning: Latent variable models for time series.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: alternative algorithms for category learning. *Psychological review*, *117*(4), 1144.

Sanders, H., Wilson, M. A., & Gershman, S. J. (2020). Hippocampal remapping as hidden state inference. *Elife*, *9*, e51140.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, *55*(4), 189.

Tomov, M. S., Dorfman, H. M., & Gershman, S. J. (2018). Neural computations underlying causal structure learning. *Journal of Neuroscience*, *38*(32), 7143–7157.

Tomov, M. S., Tsividis, P. A., Pouncy, T., Tenenbaum, J. B., & Gershman, S. J. (2023). The neural architecture of theory-based reinforcement learning. *Neuron*, *111*(8), 1331–1344.

Ullman, T. D., Goodman, N. D., & Tenenbaum, J. B. (2012). Theory learning as stochastic search in the language of thought. *Cognitive Development*, *27*(4), 455–480.

Wiltschko, A. B., Johnson, M. J., Iurilli, G., Peterson, R. E., Katon, J. M., Pashkovski, S. L., ... Datta, S. R. (2015). Mapping sub-second structure in mouse behavior. *Neuron*, *88*(6), 1121–1135.

Wu, Z., Schrater, P., & Pitkow, X. (2018). Inverse rational control: Inferring what you think from how you forage. *arXiv preprint arXiv:1805.09864*.