Learning from language and experience

Mark K. Ho (mkh260@nyu.edu)
Center for Data Science, NYU
New York, United States

Todd M. Gureckis (todd.gureckis@nyu.edu)
Department of Psychology, NYU
New York, United States

Abstract

Humans learn by directly interacting with the environment, but we also learn indirectly by communicating with other people—most notably, via language. How do people combine experience and language to learn? Here, we develop a Bayesian reinforcement learning model that integrates information from interaction with a task and propositional "hints" about a task. These hints range from low-level instructions (e.g., "pull the left arm") to concrete assertions about the task (e.g., "the left arm has an average reward of 35") to abstract relational statements (e.g., "two of the arms yield similar rewards"). Using simple bandit simulations, we show how Bayesian integration of linguistic hints can be efficiently computed and how it shapes learning. Although we mainly focus on the *computational*-level problem of language/experience integration, this work provides insights into testable *algorithmic* accounts of how people reap the benefits of joint statistical-symbolic learning.

Keywords: reinforcement learning; language; Bayesian inference; symbols; decision-making

Introduction

Both language and experience are valuable sources of information, but they rely on seemingly incommensurable cognitive processes. On the one hand, language comprehension exemplifies higher-level processing—it is rapid, symbolic, and can incorporate explicit inferences about a speaker's intentions, among other factors (Goldberg, 2003; Goodman & Frank, 2016). On the other hand, experiential learning from environmental contingencies seems to lie at the opposite extreme—it is slow, statistical, and often implicit (Dayan & Niv, 2008; Sutton & Barto, 2018). Yet, people clearly learn from both language and experience, and the two can interact (e.g., someone can recommend a restaurant we've already been to which alters our valuation of it). Our goal is investigate the computational principles that make this interaction possible.

Here, we characterize joint learning from language and experience by combining Bayesian reinforcement learning (RL) (Chapelle & Li, 2011) with formal semantics (Cresswell, 2006). We start at the computational-level (Marr, 1982) by describing the general problem of integrating language and experience. Then, we propose an efficient algorithm for language/experience integration and report simulations with RL agents given linguistic "hints". In ongoing work, we are exploring different tasks and algorithmic accounts for testing language/experience integration in participants.

Model

Bayesian reinforcement learning

The simplest setting for learning is the stationary multi-armed bandit (Sutton & Barto, 2018), in which an agent is faced with a set of arms, each of which returns a reward sampled from a fixed distribution. Here, we consider N-armed Gaussian bandits with unknown means $\mu_{1:N}$ but known variances $\sigma^2_{1:N}$. On each timestep t, the decision-maker selects an arm a_t and receives a reward $r_t \sim \mathcal{N}(\mu_a, \sigma^2_a)$. The *learning history* up to time t is the full sequence of actions taken and rewards received up until that point, $h_t = (a_1, r_1, a_2, r_2, ..., a_t, r_t)$.

We model the decision-maker as a Bayesian reinforcement learning (RL) agent that maintains a posterior distribution over the parameters of the task conditioned on the learning history:

$$P(\mu_{1:N} \mid h_t) \propto P_0(\mu_{1:N}) P(h_t \mid \mu_{1:N})$$

and selects actions using Thompson sampling (Thompson, 1933; Wilson, Bonawitz, Costa, & Ebitz, 2021).

Formalizing the semantics of task hints

Following related work on formal semantics and Bayesian models of language (Cresswell, 2006; Goodman & Frank, 2016), we treat the meanings of linguistic hints as functions over possible parameterizations of a task Θ (here, $\Theta = \mathbb{R}^N$, the space of all N-arm Gaussian bandit configurations). Formally, the meaning of a linguistic *hint* l is a function from task parameterizations to real numbers extended with negative infinity, $f_l : \Theta \to \mathbb{R} \cup \{-\infty\}$. In the context of Gaussian bandits, our use of (extended) real-valued functions allows us to capture hints that are categorically true or false, like "Arm 1 is more than 25" (falsehood is represented as $-\infty$; truth as 0), as well as graded similarity/difference relationships between parameter values, such as "Arm 1 is similar to arm 2" in a manner analogous to fuzzy logics used in control engineering (Zadeh, 1965).

Meaning functions are represented as arithmetic/logical *terms* that are evaluated with respect to a particular set of parameter values $\mu_{1:n} \in \Theta$ (Table 1). Terms adhere to a formal grammar of arithmetic/logical composition, similar to those used in models of rule induction (Goodman, Tenenbaum, Feldman, & Griffiths, 2008), but here, we have hand-encoded hints to their terms. For example, we encode the hint "Arm 1 is more than 25" as $\ln \mathbf{1}(\mu_1 > 25)$, which includes a Boolean subterm ($\mu_1 > 25$), an indicator function ($\mathbf{1}(\cdot)$), and the natural log function ($\ln(\cdot)$).

The grammar also allows us to express abstract, quantified propositions by encoding existential/universal quantification as the *max*/*min* value over subterms with variables, similar to how disjunction/conjunction are encoded in fuzzy logics (Zadeh, 1965). For instance, the hint "One of the arms is greater than 25" would be encoded as $\max_i \{\ln \mathbf{1}(\mu_i > 25)\}$—this term contains a variable for arms i that allows us to encode abstract propositions about the set of arms.

Table 1: Example hint encodings

Linguistic Hint (l)	Meaning function (f_l)
"Arm 1 is more than 25"	$\ln \mathbf{1}(\mu_1 > 25)$
"Arm 1 is around 40"	$-\lvert\mu_1 - 40\rvert$
"Arm 1 is similar to arm 2"	$-\lvert\mu_1 - \mu_2\rvert$
"Arm 3 is the best"	$\ln \mathbf{1}(\arg\max_a \mu_a = 3)$
"One of the arms is more than 25"	$\max_a \{\ln \mathbf{1}(\mu_a > 25)\}$
"Two distinct arms are similar"	$\max_{i,j:i\neq j} \{-\lvert\mu_i - \mu_j\rvert\}$

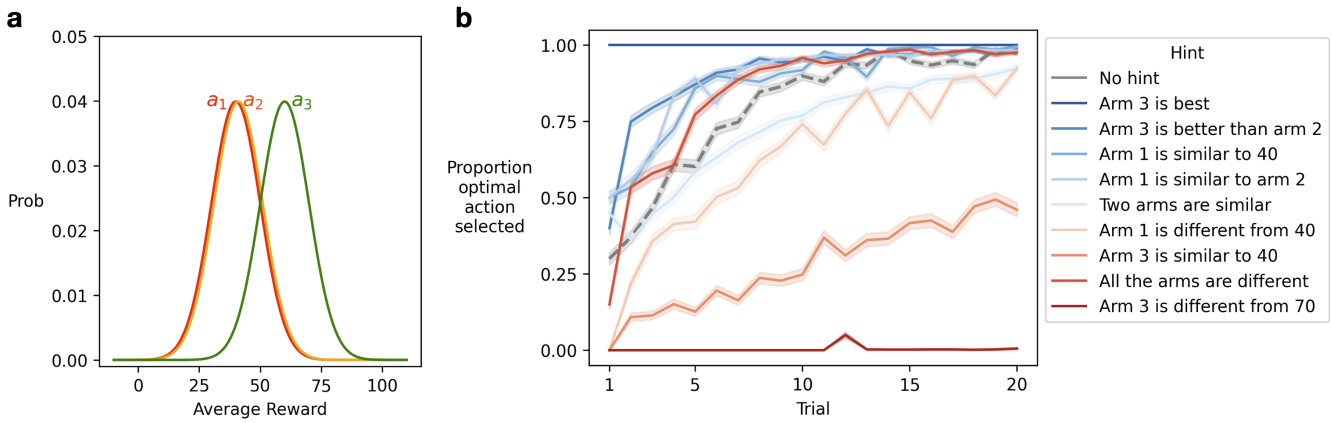

Figure 1: Simulated learning with different hints. (a) Bayesian RL agents were trained for 20 trials on a three-armed Gaussian bandit, $\mu_1 = 40$, $\mu_2 = 41$, $\mu_3 = 60$. For all arms, $\sigma_a = 10$. (b) Learners were trained with different hints that were helpful (blue) or misleading (red). Hints had different effects on learning relative to an agent given no hint (grey). Plotted are the proportion of times the optimal action (a_3) was selected on each trial. For each hint, 20 agents were trained on 100 runs on the bandit with matching pseudo-random seeds. Error bands are 95% CI.

Integrating language and experience

Bayesian approaches provide a general framework for integrating information from different sources as posterior inference. This includes joint learning from language and experience in a Gaussian bandit. Formally, given a linguistic hint l with meaning function f_l and learning history h_t, the posterior beliefs over mean arm rewards $\mu_{1:N}$ is:

$$P(\mu_{1:N} \mid h_t, l) \propto \underbrace{P_0(\mu_{1:N})}_{\text{Prior}} \underbrace{P(h_t \mid \mu_{1:N})}_{\text{Experience}} \underbrace{\exp\{f_l(\mu_{1:N})\}}_{\text{Hint}}. \quad (1)$$

How can an RL agent efficiently compute or even approximate Equation 1 at every time t? In general, updating an arbitrary prior with an arbitrary likelihood is intractable, however, given certain assumptions about the representation of these components, inference can be made efficient. For example, Table 2 shows pseudo-code for an algorithm that efficiently integrates task feedback with a hint to approximate Equation 1. Conceptually, inference is accomplished in two stages: first, a generative model that only takes into account the learning history h_t is updated analytically using conjugate priors (Murphy, 2022). Second, K samples are taken from the experience-only posterior and filtered by calculating the weights according to the hint's meaning function f_l (Shachter & Peot, 1990). The normalized weights then serve as approximations to the probability of each sample conditioned on the hint.

Using the algorithm sketched out in Table 2, we ran a series of simulations in which Bayesian RL agents were given no hint, informative hints, or misleading hints. As shown in Figure 1, different hints shaped learning in a variety of ways.

Table 2: Language/experience integration algorithm

Input: Learning history h_t, hint function f_l, arm prior hyperparams $\bar{\mu}$, $\bar{\sigma}^2$, arm variances $\sigma^2_{1:N}$, particle count K
Output: Empirical posterior $\hat{P}(\mu_{1:N} \mid h_t, l)$
for $a \in [N]$ **do** `//Update model with experience`
  $N_a = \sum_{(a_t, r_t) \in h_t} \mathbf{1}(a_t = a)$ `//times a was pulled`
  $R_a = \sum_{(a_t, r_t) \in h_t} r_t \mathbf{1}(a_t = a)$ `//total reward from a`
  $\bar{\sigma}^{2\prime}_a = \left(\frac{1}{\bar{\sigma}^2} + \frac{N_a}{\sigma^2_a}\right)^{-1}$ `//updated variance of` μ_a
  $\bar{\mu}'_a = \left(\frac{\bar{\mu}}{\bar{\sigma}^2} + \frac{R_a}{\sigma^2_a}\right) \bar{\sigma}^{2\prime}_a$ `//updated mean of` μ_a
end for
for $i \in [K]$ **do** `//Filter with hint`
  $\mu^{(i)}_{1:N} \sim \mathcal{N}(\bar{\mu}_{1:N}, \bar{\sigma}^2_{1:N})$ `//sample arm means`
  $w^{(i)} = f_l(\mu^{(i)}_{1:N})$ `//calculate weight`
end for
$\hat{P}(\mu^{(i)}_{1:N} \mid h_t, l) \propto \exp\{w^{(i)}\}$ `//normalize weights`
return $\hat{P}(\mu^{(i)}_{1:N} \mid h_t, l)$

Discussion

Here, we have formulated the problem of optimally integrating linguistic hints and task feedback, demonstrated how integration can performed efficiently, and investigated how hints can shape learning in a simple bandit task. This approach complements previous work on how instructions shape human RL from a neurocomputational perspective (e.g., Doll, Jacobs, Sanfey, & Frank, 2009). Additionally, our proposed algorithm provides a starting point for testing different mechanistic theories of language/experience integration in humans. For example, although here we assume that language mainly influences "filtering", hints could also shape how people "generate" samples of task parameters to begin with (e.g., when an experimenter instructs participants to view the arms as independent Gaussians, this induces a particular generative model). Future work will investigate the different ways in which symbolic communication in the form of language interacts with experiential learning by comparing the predictions of these algorithms with human behavior.

References

Chapelle, O., & Li, L. (2011). An empirical evaluation of thompson sampling. *Advances in neural information processing systems*, *24*.

Cresswell, M. (2006). Formal semantics. *The Blackwell guide to the philosophy of language*, 131–46.

Dayan, P., & Niv, Y. (2008). Reinforcement learning: the good, the bad and the ugly. *Current opinion in neurobiology*, *18*(2), 185–196.

Doll, B. B., Jacobs, W. J., Sanfey, A. G., & Frank, M. J. (2009). Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. *Brain research*, *1299*, 74–94.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in cognitive sciences*, *7*(5), 219–224.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, *20*(11), 818–829.

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.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman and Company.

Murphy, K. P. (2022). *Probabilistic machine learning: An introduction*. MIT Press. Retrieved from `probml.ai`

Shachter, R. D., & Peot, M. A. (1990). Simulation approaches to general probabilistic inference on belief networks. In *Machine intelligence and pattern recognition* (Vol. 10, pp. 221–231). Elsevier.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, *25*(3/4), 285–294.

Wilson, R. C., Bonawitz, E., Costa, V. D., & Ebitz, R. B. (2021). Balancing exploration and exploitation with information and randomization. *Current opinion in behavioral sciences*, *38*, 49–56.

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, *8*(3), 338–353.