Journal of Mathematical Psychology 77 (2017) 10–20

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Online learning of symbolic concepts

Pratiksha Thaker [a], Joshua B. Tenenbaum [b], Samuel J. Gershman [c,*]

[a] Department of Computer Science, Stanford University, United States
[b] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, United States
[c] Department of Psychology and Center for Brain Science, Harvard University, United States

HIGHLIGHTS

- A novel variant of the number game is studied.
- An online approximate Bayesian model captures order effects in concept learning.
- Placing people under cognitive load strengthens the order effect.

ARTICLE INFO

Article history:
Received 30 July 2016
Received in revised form
28 November 2016

Keywords:
Bayesian models
Concept learning
Bounded rationality
Particle filtering

ABSTRACT

Learning complex symbolic concepts requires a rich hypothesis space, but exploring such spaces is intractable. We describe how sampling algorithms can be brought to bear on this problem, leading to the prediction that humans will exhibit the same failure modes as sampling algorithms. In particular, we show that humans get stuck in "garden paths"—initially promising hypotheses that turn out to be sub-optimal in light of subsequent data. Susceptibility to garden paths is sensitive to the availability of cognitive resources. These phenomena are well-explained by a Bayesian model in which humans stochastically update a sample-based representation of the posterior over a compositional hypothesis space. Our model provides a framework for understanding "bounded rationality" in symbolic concept learning.

1. Introduction

One of the most remarkable characteristics of human cognition is the ability to learn symbolic concepts from very sparse data. For example, after being shown the numbers {60, 80, 10, 30} drawn from a set of numbers between 1 and 100, humans will confidently infer the set to be "multiples of ten" (Tenenbaum, 1999; Tenenbaum & Griffiths, 2001). This kind of strong generalization requires a hypothesis space rich enough to express a wide variety of concepts, as well as a mechanism for efficiently exploring the hypothesis space and evaluating candidate concepts. A conundrum at the heart of concept learning is that these two requirements are at odds with one another: The richer the hypothesis space, the harder it is to efficiently explore. This is especially true for compositional hypothesis spaces (e.g., Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Kemp, 2012; Piantadosi, Tenenbaum, & Goodman, 2010), where the number of possible concepts is exponential in the number of primitives. Moreover, Bayesian approaches to concept learning assert that humans represent a probability distribution over the entire hypothesis space (Shepard, 1987; Tenenbaum, 1999; Tenenbaum & Griffiths, 2001). These considerations bring the issue of computational tractability to the foreground.

Previous treatments of symbolic concept learning have primarily focused either on abstract rational analysis without detailed mechanistic commitments (Feldman, 2000; Kemp, 2012; Piantadosi et al., 2010; Tenenbaum, 1999; Tenenbaum & Griffiths, 2001) or on mechanistic models without a clear connection to rational inductive principles (Goodwin & Johnson-Laird, 2011; Kruschke et al., 1992; Nosofsky, Palmeri, & McKinley, 1994). Goodman et al. (2008) used a compositional grammar to model boolean concept learning, and presented provisional evidence that participants adhere to rational inductive principles only approximately: Behavior was best explained by assuming that humans make their responses using one or a few samples from the posterior distribution over concepts. Hypothesis sampling has become an important bridge between rational analyses and process models (see Griffiths, Vul, & Sanborn, 2012, for a review), with applications to vision (Gershman, Vul, & Tenenbaum, 2012; Moreno-Bote, Knill, & Pouget,

* Correspondence to: Department of Psychology, Harvard University, 52 Oxford St., room 295.05, Cambridge, MA 02138, United States.
E-mail address: gershman@fas.harvard.edu (S.J. Gershman).

http://dx.doi.org/10.1016/j.jmp.2017.01.002
0022-2496/

2011; Vul, Frank, Alvarez, & Tenenbaum, 2009; Wozny, Beierholm, & Shams, 2010), theory learning (Denison, Bonawitz, Gopnik, & Griffiths, 2013; Ullman, Goodman, & Tenenbaum, 2012) and categorization (Sanborn, Griffiths, & Navarro, 2010), among others. Empirical evidence for hypothesis sampling will be reviewed in the General Discussion.

Hypothesis sampling provides a simple model of cognitive limitations (in terms of how many samples are used) while instantiating a theoretically sound mechanism for approximating Bayesian inference (Vul, Goodman, Griffiths, & Tenenbaum, 2014). In particular, many hypothesis sampling models can be viewed as Monte Carlo methods, which are widely used in statistics and machine learning due to their flexibility and theoretical properties (Robert & Casella, 2004). Previous work on concept learning in compositional hypothesis spaces used Markov chain Monte Carlo (MCMC) algorithms to generate samples (Goodman et al., 2008; Piantadosi et al., 2010); since these algorithms evaluate hypotheses over the entire data set at each iteration, they are cognitively implausible for tasks in which data are presented sequentially and presumably processed online (as in many domains, like word learning or multiple-object tracking).

This paper investigates a cognitively plausible sampling algorithm for performing online inference over a compositional hypothesis space of number concepts. Our starting point is the "number game" described in Tenenbaum (1999). In this experiment, participants were presented with a set of integers generated from a number concept (a subset of numbers between 1 and 100), such as "all powers of 2" or "all numbers between 40 and 60". Participants were then asked to judge, for several other numbers, the probability that each was generated from the same subset as the examples presented. Tenenbaum (1999) showed that generalization patterns in this experiment were consistent with a Bayesian model of concept learning (described in more detail below). While the space of number concepts is very large, Tenenbaum's model constrained the hypothesis space to a small number of intuitively plausible concepts.

We will consider a richer space of compositional concepts, and postulate a form of hypothesis sampling as a theory of how humans explore this hypothesis space. In particular, we argue that humans use an online hypothesis sampling algorithm called *particle filtering* that entertains multiple hypotheses ("particles") and continually reweights the particles as new data are observed. This algorithm has previously been used to explain aspects of multiple object tracking (Vul et al., 2009), category learning (Sanborn et al., 2010), change detection (Brown & Steyvers, 2009), word segmentation (Frank, Goldwater, Griffiths, & Tenenbaum, 2010), and reinforcement learning (Daw & Courville, 2007; Yi, Steyvers, & Lee, 2009). While most of this previous work has focused on hypothesis spaces with relatively simple representational structure (e.g., mixture models), our goal is to provide empirical constraints on hypothesis sampling in more complex symbolic spaces.

One implication of hypothesis sampling is that when faced with complex or ambiguous example sets in the number game, participants might fail to infer some concepts that have high posterior probability. We speculated that this might happen if examples are presented to participants sequentially, such that the early examples favor one concept, but the later examples tilt the posterior in favor of a different concept. If conditionally unlikely samples are eliminated during hypothesis sampling (an operation known as "resampling"), the early, sub-optimal concept will prevail. This is analogous to "garden path" sentences in psycholinguistics (e.g., "we painted the walls with cracks") which are difficult for humans to parse (MacDonald, 1994). Levy, Reali, and Griffiths (2009) modeled garden path effects with hypothesis sampling by assuming that the correct parse was eliminated from the sample set early on during sentence processing. We adapted this model to number concept learning, and constructed example sequences which would lead the model to show garden path effects. We then conducted experiments with humans to test whether participants show the same effects.

The plan of the paper is as follows. Section 2 summarizes the Bayesian framework for concept learning developed by Tenenbaum (1999), and introduces a hypothesis sampling algorithm for approximate inference. Section 3 reports an experiment with human participants playing a sequential concept learning game. We show how the hypothesis sampling algorithm provides a rational process account of order effects and cognitive load manipulations in the game. Section 4 concludes the paper with a discussion of related work and future directions.

2. A Bayesian framework for concept learning

In this section, we describe and extend the Bayesian framework for concept learning introduced by Tenenbaum (1999). We begin by describing the *generative model*—a joint distribution over concepts and data. The generative model specifies the learner's assumptions about what types of concepts are plausible (the prior) and how concepts give rise to observations (the likelihood). Of central importance is our claim about concept representation: Number concepts are sets generated by a compositional, probabilistic grammar. We then describe how hypothesis sampling can be used to perform approximate inference over number concepts. This sampling-based rational process model provides the basis for our experimental investigations.

Before proceeding, we provide here a non-technical summary of how the model works. A concept is drawn from some space of plausible concepts (the hypothesis space), and examples are drawn from the selected concept. The learner's job is to infer the hidden concept that generated the examples. Because many different concepts can generate any particular set of examples, the problem is fundamentally ill-posed: No single concept is unambiguously "correct". Rather, the optimal inductive inference is a *distribution* over concepts (the posterior), which is computed by multiplying the prior and the likelihood for each potential concept, and then normalizing over the hypothesis space. However, a combinatorial hypothesis space may contain too many hypotheses for complete enumeration to be tractable. A solution to this problem is to randomly sample hypotheses from the posterior and approximate the distribution with a histogram—this is the basis of *Monte Carlo methods* (Robert & Casella, 2004). By limiting the number of samples, a learner can trade off cognitive resources with accuracy: A larger number of samples consumes more cognitive resources (in terms of memory and processing time) while producing a more accurate approximation of the posterior. As we show experimentally, reducing the availability of cognitive resources has deleterious effects on the accuracy of the posterior.

One challenge for practical applications of Monte Carlo methods is that we cannot easily sample from the posterior. To surmount this challenge, we can instead sample from a proposal distribution (e.g., the prior) and then weight the samples to correct for the fact that they were generated from the wrong distribution. When the number of samples is small and the proposal distribution is far from the posterior, this method can lead to degeneracy: a small number of samples have very large weights and the rest of the samples are effectively ignored. This means that the *effective* sample size is smaller than the number of samples. To remedy this problem, we can delete conditionally unlikely samples (i.e., those with small weights) by resampling: generating a new sample set by drawing samples with probability proportional to their weights.

A final challenge is that the examples may arrive sequentially, and it is wasteful to recompute the posterior from scratch after