Cognition 168 (2017) 320–334

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Learning abstract visual concepts via probabilistic program induction in a Language of Thought ☆

Matthew C. Overlan, Robert A. Jacobs *, Steven T. Piantadosi

Department of Brain & Cognitive Sciences, University of Rochester, Rochester, NY 14627, United States

A R T I C L E   I N F O

Article history:
Received 14 February 2017
Revised 6 July 2017
Accepted 9 July 2017
Available online 1 August 2017

Keywords:
Concept learning
Visual learning
Language of Thought
Computational modeling
Behavioral experiment

A B S T R A C T

The ability to learn abstract concepts is a powerful component of human cognition. It has been argued that variable binding is the key element enabling this ability, but the computational aspects of variable binding remain poorly understood. Here, we address this shortcoming by formalizing the Hierarchical Language of Thought (HLOT) model of rule learning. Given a set of data items, the model uses Bayesian inference to infer a probability distribution over stochastic programs that implement variable binding. Because the model makes use of symbolic variables as well as Bayesian inference and programs with stochastic primitives, it combines many of the advantages of both symbolic and statistical approaches to cognitive modeling. To evaluate the model, we conducted an experiment in which human subjects viewed training items and then judged which test items belong to the same concept as the training items. We found that the HLOT model provides a close match to human generalization patterns, significantly outperforming two variants of the Generalized Context Model, one variant based on string similarity and the other based on visual similarity using features from a deep convolutional neural network. Additional results suggest that variable binding happens automatically, implying that binding operations do not add complexity to peoples' hypothesized rules. Overall, this work demonstrates that a cognitive model combining symbolic variables with Bayesian inference and stochastic program primitives provides a new perspective for understanding people's patterns of generalization.

1. Introduction

Induction, the ability to discover latent patterns and structure from a set of data items, is a hallmark of human thinking. This ability underlies our remarkable language acquisition and conceptual development, and its roots have been found in infancy. Marcus, Vijayan, Bandi Rao, and Vishton (1999) studied the ability of seven-month-olds to infer abstract rules from acoustic sequences. They showed that infants presented with syllable sequences that follow an *ABA* pattern, like "ga ti ga" and "li na li", recognized novel sequences following that pattern even when those sequences contained new syllables that the infants had not heard, like "wo fe wo". Because the test items could not be distinguished based on concrete features like transitional statistics between syllables, sequence length, or prosody, they reasoned that the infants had learned an abstract rule that reflected latent structure.

Even though rules like *ABA* are simple, they illustrate a foundational computational element of human abstract rule learning: we can easily and fluidly handle *variable binding* (Jackendoff, 2003; Marcus, 2003). Variable binding refers to the ability to assign a name to some piece of information for storage and later retrieval. In the case of *ABA* rules, infants must remember the first syllable (that is, store it in a variable *A*) so it can be compared to subsequent syllables. The use of variables is what allows the *ABA* rule to be abstract: computations can refer to variable names rather than the values stored therein, so the rule can reflect the *relationship* between pieces of information rather than the concrete features of that information. It does not matter what values the *A*s and *B*s have, so long as the resulting sequence obeys the right pattern of repetition.

Variable binding has been at the center of a key debate in cognitive science (Jackendoff, 2003; Marcus, 2003), much of which has focused on the role of statistics in learning abstract rules.

☆ A preliminary version of a portion of this research appeared in the *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*. We thank Goker Erdogan and Frank Mollica for helpful discussions, and Goker Erdogan for neural network code. This work was supported by research grants from the Air Force Office of Scientific Research (FA9550-12-1-0303) and the National Science Foundation (BCS-1400784).

* Corresponding author.
E-mail addresses: m.overlan@rochester.edu (M.C. Overlan), robbie@bcs.rochester.edu (R.A. Jacobs), spiantadosi@bcs.rochester.edu (S.T. Piantadosi).

http://dx.doi.org/10.1016/j.cognition.2017.07.005
0010-0277/© 2017 Elsevier B.V. All rights reserved.

Proponents of a rule-based approach point out that statistics alone are insufficient for learning rules that require variable binding, since (as with *ABA* rules) variable binding allows learners to generalize to novel stimuli for which they have no statistical information.[1] In response, proponents of a statistical approach point out that a pure rule-based approach cannot explain why learners choose the rules they do from the infinitely many that are consistent with the input. In addition, infant studies have shown that the statistics of the input affect generalization of *ABA*-like rules. For instance, Gerken (2006) presented infants with syllable sequences that were logically consistent with two different rules and showed that they learned the one that was best supported by the statistics of the input they had received.

This tension between rules and statistics has been addressed in recent years by hybrid models, sometimes referred to as probabilistic language of thought (pLOT) models (Piantadosi & Jacobs, 2016). pLOT models operate with infinite hypothesis spaces by employing a compositional system for creating rules (Erdogan, Yildirim, & Jacobs, 2015; Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Kemp, 2012; Siskind, 1996; Piantadosi, Tenenbaum, & Goodman, 2012; Piantadosi, Tenenbaum, & Goodman, 2016; Ullman, Goodman, & Tenenbaum, 2012; Yildirim & Jacobs, 2015). These models integrate rules and statistics by employing statistical (e.g., Bayesian) inference over such structured hypothesis spaces (Tenenbaum & Griffiths, 2001). By using structured, symbolic hypotheses, these models can represent "rule-based" concepts. And by maintaining uncertainty over rules, these models can operate in the presence of noisy data showing gradience or typicality effects.

While there have been several process-level models of variable binding in neural networks (Hummel & Holyoak, 1997; Smolensky, 1990), few models have approached the problem of variable binding from the ideal-observer (Geisler, 2003) perspective, considering a computational-level explanation for rule learning as the rational outcome of an optimal computation.

Frank and Tenenbaum (2011) implemented an ideal-observer model of *ABA*-style rule learning in which variables are *implicit*. They represented these rules with 3-tuples like $(is_{ga}, *, =_1)$, which meant that the first syllable is a specific one ('ga' in this case), the second syllable is free (it could be anything), and the third syllable is equal to the first. Their model used Bayesian techniques to approximately capture generalization patterns by performing inference over this space of hypotheses. While this representation is strictly sufficient to capture *ABA*-like patterns, it has important shortcomings. Since its variables are simply built in as a baseline in the representation, their model is unable to explain why learners may or may not come to hypothesize variables in the first place. Second, the space of concepts and rules lacks any prior biases over hypotheses. In particular, there is no notion of simplicity or complexity, a key inductive bias for human learners (Chater & Vitanyi, 2003; Feldman, 2000). A simplicity bias allows learners to avoid over-fitting and to come to a reasoned compromise between generality and fit-to-data. Finally, their representation is highly specific to identity-based, *ABA*-like patterns. This makes it unclear how their methods and ideas might generalize to the many other classes of rule-based concepts that have been studied in the literature.

Our model addresses all of these issues. We represent concepts as *probabilistic programs*, programs with stochastic primitives such that they produce different random outputs each time they are run. A program-based representation allows hypotheses in our model to contain explicit variable binding operations. To infer these programs from data, we build upon the pLOT framework's capability of Bayesian statistical inference over a structured space of symbolic hypotheses. Following Goodman et al. (2008), we assume a rich generative model for concepts that uses a probabilistic context-free grammar to represent an infinite space of hypotheses. This grammar-based approach provides a natural simplicity-favoring prior over programs. These qualities of explicit variable binding and robust statistical inference allow us to reason about abstraction in rule learning in a way that is not possible with a fixed assignment of items to slots and a uniform prior.

At the highest level of generality, the goal of our research program is to characterize human learning and reasoning as forms of program induction. We regard the pLOT as a promising framework for developing such a characterization. Our model combines a symbolic approach, which provides a means for achieving abstraction (through variable binding) and for defining an infinite structured hypothesis space (through compositionality), with a statistical approach, which provides a means for learning representations from noisy data in a way that quantifies uncertainty (through Bayesian inference and the use of programs with stochastic primitives). A novel innovation of our model is that it combines statistical program induction (i.e., Bayesian inference of a probability distribution over programs) with the use of probabilistic programs (i.e., those with stochastic primitives). We see the work presented in this paper as an early step toward extending symbolic-statistical hybrid models so that they can be used to develop theoretical accounts in many domains of human cognition.

In addition to the theoretical contributions of our computational framework, our secondary goal is to provide empirical results that further our understanding of human rule learning. Currently, our knowledge of human learning of *ABA*-like rules is limited to data available from infant studies. The necessary sparseness of these data makes it difficult to distinguish between competing models at a fine grain. Therefore, we carried out a behavioral experiment with adults that is inspired by infants' learning of *ABA*-like patterns. This allows us to assess subjects' generalization patterns at a detailed level.

The plan of this paper is as follows. First, we describe the details of our behavioral experiment, then the details of our model. We next compare the generalization performance of our model with those of our experimental subjects. We find that our model provides an excellent account of our experimental data, outperforming alternative models that lack key elements such as variable binding. Finally, we test a variant of our model in order to determine which way of handling variable abstractions provides the most accurate fit to human generalizations.

2. Behavioral experiment

In our behavioral experiment, we evaluated human subjects' abilities to infer an abstract visual concept or category from a small number of exemplars. This was accomplished by showing subjects exemplars consistent with a concept, and then asking them whether they believed each of several test items was also an exemplar from the same concept. All subjects were US residents over the age of 18. They participated in the experiment over the world wide web using the Amazon Mechanical Turk crowdsourcing platform. Raw data from the experiment can be found in the online supplemental materials.

Visual stimuli were images depicting 3D, part-based objects rendered with realistic lighting and texture (see Fig. 1a for the set of possible object parts). Based on these images, it was easy to segment an object into its component parts. Our stimuli have the advantage of being both novel—meaning that subjects did

[1] While much of this debate has focused on the deficiencies of connectionist models (Fodor & Pylyshyn, 1988) and possible connectionist solutions (Gayler, 2004; Smolensky, 1990; Smolensky & Legendre, 2006; van der Velde & de Kamps, 2006), these arguments apply to any sub-symbolic theory that does not have an explicit representation of variables.