Cognitive Psychology 99 (2017) 44–79

Contents lists available at ScienceDirect

Cognitive Psychology

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

Compositional inductive biases in function learning

Eric Schulz[a,*], Joshua B. Tenenbaum[b], David Duvenaud[c], Maarten Speekenbrink[d], Samuel J. Gershman[a]

[a] Harvard University, United States
[b] Massachusetts Institute of Technology, United States
[c] University of Toronto, Canada
[d] University College London, United Kingdom

ARTICLE INFO

Keywords:
Function learning
Pattern recognition
Compositionality
Structure search
Gaussian process

ABSTRACT

How do people recognize and learn about complex functional structure? Taking inspiration from other areas of cognitive science, we propose that this is achieved by harnessing compositionality: complex structure is decomposed into simpler building blocks. We formalize this idea within the framework of Bayesian regression using a grammar over Gaussian process kernels, and compare this approach with other structure learning approaches. Participants consistently chose compositional (over non-compositional) extrapolations and interpolations of functions. Experiments designed to elicit priors over functional patterns revealed an inductive bias for compositional structure. Compositional functions were perceived as subjectively more predictable than non-compositional functions, and exhibited other signatures of predictability, such as enhanced memorability and reduced numerosity. Taken together, these results support the view that the human intuitive theory of functions is inherently compositional.

1. Introduction

Recognizing functional patterns is a ubiquitous problem in everyday cognition, underlying the perception of time, space and number. How much food should you cook to satisfy every guest at a party? How far do you have to turn the faucet handle to get the right temperature? Should you invest in a particular stock that seems to be going up? Since the space of such mappings is theoretically infinite, inductive biases are necessary to constrain the plausible inferences (Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; Mitchell, 1980). But what is the nature of these inductive biases over functions?

The major theoretical accounts of function learning, which we review below, offer two different answers to this question. Rule-based accounts posit a fixed set of parametric forms (or rules) that serve as a "vocabulary" for functions; these accounts imply strong inductive biases for the rules in the functional vocabulary. By contrast, similarity-based accounts posit a nonparametric representation of functions, implying relatively weak inductive biases.

A major challenge for humans is how to accommodate the virtually infinite diversity of functions in the real world. Rule-based models can only represent linear combinations of a fixed set of parametric functions. Similarity-based models can in principle represent an infinite variety of functions, but their typically weak inductive biases do not support strong inferences from small amounts of data—an important characteristic of human learning (Lake, Ullman, Tenenbaum, & Gershman, 2016).

A ubiquitous strategy in many areas of cognition, from language (Chomsky, 1965) to concept learning (Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Kemp, 2012; Piantadosi, Tenenbaum, & Goodman, 2016) and visual perception (Biederman, 1987; Lake,

* Corresponding author at: Department of Psychology, Harvard University, 52 Oxford Street, Room 295.08, Cambridge, MA 02138, United States.
E-mail address: ericschulz@fas.harvard.edu (E. Schulz).

https://doi.org/10.1016/j.cogpsych.2017.11.002
Accepted 6 November 2017

Salakhutdinov, & Tenenbaum, 2015), is to divide and conquer: construct complex representations out of simpler building blocks using a set of compositional rules. Compositional systems support strong inferences from small amounts of data by imposing structural constraints, without sacrificing the capacity for representing an infinite variety of forms. The primary claim in this paper is that human function learning is structurally constrained by compositional inductive biases.

To formalize this idea, we need a theoretical framework for function learning that can represent and reason about compositional function spaces. Lucas, Griffiths, Williams, and Kalish (2015) recently presented a normative theory of function learning using the formalism of Gaussian processes (GPs). As we will describe more formally, GPs are distributions over functions that can encode properties such as smoothness, linearity, periodicity, symmetry, and many other inductive biases found by past research on human function learning (Brehmer, 1974b; DeLosh, Busemeyer, & McDaniel, 1997). Lucas et al. (2015) showed how Bayesian inference with GP priors can be expressed in both parametric (rule-based) and nonparametric (similarity-based) forms. GPs can therefore serve as a computational-level theory of function learning that bridges different mechanistic implementations.

In this paper, we build on the GP formalism to study, both theoretically and experimentally, the compositional nature of inductive biases in human function learning. Our extensions of the GP formalism not only bridge the "rules" and "similarity" perspectives on learning, but can also explain how people are able to learn much more complex kinds of functional relationships that are not well described by either traditional notions of rules or traditional kinds of similarity metrics.

Our main theoretical contribution is to extend the GP approach to modeling human function learning with a prior that obeys compositionally structured constraints. We do this using a compositional grammar for intuitive functions introduced in the machine learning literature by Duvenaud, Lloyd, Grosse, Tenenbaum, and Ghahramani (2013). We then test the predictive and explanatory power of this compositional GP model in 10 function learning and reasoning experiments, comparing the compositional prior to a flexible non-compositional prior (the spectral mixture representation proposed by Wilson & Adams (2013), which we will describe later). Both models use Bayesian inference to reason about functions, but differ in their inductive biases.

Our experiments begin by comparing these different models of human function learning on five functional pattern completion tasks, two of which ask participants to choose among different completions, two of which assess a restricted posterior distribution over compositional kernels, and one of which asks participants to manually complete sampled functions within a graphical user interface.

Throughout all of these experiments, we find that participants' completions are better described by the compositional prior as compared to non-compositional alternatives. We then generate a set of 40 similar functions, 20 of which are compositional and 20 of which are non-compositional, and compare these functions by asking participants how predictable they are and letting them learn and predict these functions in two experiments using a trial-by-trial function learning paradigm. We find that participants not only perceive compositional functions as more predictable and learn them more easily, but also that a compositional model of both predictability and function learning provides a quantitatively accurate description of participants' behavior. Finally, we investigate how compositional functions influence memory, change detection, and the perception of numerosity in three additional experiments. To understand these results computationally, we propose a compositional Bayesian model of pattern encoding, chunking and retrieval, and compare this model to other alternatives. We conclude by discussing the implications as well as possible limitations of our proposed model and spell out future directions of compositional function learning research.

2. Prior research on human function learning

The general problem of inferring how one variable depends functionally on another is important for many aspects of cognition. Traditionally, it has been studied in paradigms assessing how people learn about input-output mappings or how they make sense of spatiotemporal patterns. Additionally, the way in which we learn about and recognize functional patterns is also crucial in the modern world as we look at data – either as scientists or non-scientist decision makers – in trying to understand what function the data reveal. We are interested in all of these aspects of function learning, but will focus first on the latter because it can potentially reveal how people recognize and perform inference about structure very quickly. However, we believe that a strength of our account of compositional inductive biases is that it can account for empirical effects across a diverse set of paradigms.

Donald Broadbent was among the first psychologists to investigate how people learn and control functions between inputs and outputs (Broadbent, 1958). In his experiments, participants controlled functions within an industrial setting called the "sugar factory," in which they learned the relationship between work force and sugar production. Broadbent showed that participants had difficulty controlling some functions, such as exponential or power functions, but were good at learning others, for example linear functions (Berry & Broadbent, 1984).

Since Broadbent's pioneering work, further studies have established several empirical regularities (see McDaniel & Busemeyer, 2005, for a review). For example, studies using interpolation judgments—predictions of function outputs for inputs inside the convex hull of training inputs—have found that linear, increasing functions are easier to learn than non-linear, non-monotonic or decreasing functions (Brehmer, 1974a; Brehmer, Alm, & Warg, 1985; Byun, 1995). Presentation order also matters: it is easier to learn functions if the input is ordered by increasing output (DeLosh et al., 1997).

Important constraints on theories of function learning have come from studies of extrapolation judgments—predictions of function outputs for inputs outside the convex hull of training inputs (DeLosh et al., 1997; McDaniel & Busemeyer, 2005). People tend to make linear extrapolations with a positive slope and an intercept of zero (Kalish, Lewandowsky, & Kruschke, 2004; Kwantes & Neal, 2006). This linearity bias holds true even when the underlying function is non-linear; for example, when trained on a quadratic function, average predictions fall between the true function and straight lines fitted to the closest training points (DeLosh et al., 1997).