Concepts and Categories

A computational perspective on constructivism

Eli Sennesh

December 11, 2018

Psychologists, Engineers, and Neuroscientists – Northeastern University

Table of contents

Introduction: Concepts in Psychology and Neuroscience

What are we talking about?

"Concept learning": one of the oldest tasks studied in cognitive psychology and machine learning. Understandings have diverged across fields.

- Psychologists talk about using concepts to build propositions, *productively*.
- Neuroscientists talk about concepts in terms of embodied simulation, re-enactment, or reinstatement.
- Machine learners talk about "concept learning" as learning a *classifier function* over an arbitrary *feature space*.

Categories as per psychologists

"In psychology, a category is a group of instances sharing a functional similarity within a context (e.g., [7])."

Questions:

- What are instances?
- What sort of functional similarity?
- What supplies the context? What does it mean to be within a context?

Intuitively "obvious" answers often need a lot of work to reduce to lower-level science. These are good targets for investigation.

Concepts as per neuroscience

“A simulator is a distributed collection of modality-specific memories captured across a category’s instances. When the category is processed on a given occasion, only a small subset of this information becomes active – not the entire simulator. The active subset is then run as a simulation that functions as one of infinitely many conceptualizations for the category.” [2]

Questions:

- What are instances?
- What matches simulators to categories of instances?
- What is an occasion?
- How can a finite brain support infinitely many conceptualizations?

What does the brain use them for?

Why bother having concepts and categories? Many AI models get by without them! Some common answers:

- Because they objectively exist as Platonic Forms,
- Because we evolved with them,
- To model the world's causal structure,
- To make decisions between one thing and another,
- To regulate interoceptive sensations,

"The difficulty of defining concept raises the issue of whether it is a useful scientific construct. Perhaps no discrete entity or event constitutes a concept."

Concepts and Categories in Machine Learning

Mathematical setup

In machine learning, we give precise meanings to some of the terms we have discussed.

- Feature space: a mathematical space of all the quantities we measure in each instance, usually $\mathbb{R}^{\mathcal{D}}$.
- Instance: a point in the feature space $x \in \mathbb{R}^{\mathcal{D}}$, or a collection of measurements. Usually sampled from the real world.
- Classifier: a function from a feature space to a yes-or-no decision, $f : \mathbb{R}^{\mathcal{D}} \to \{+1, -1\}$.

Why machine learning rather than "just" statistics?

Machine learning investigators usually work under a *distribution-free assumption*: we receive data $x_1, \ldots, x_k \in \mathbb{R}^{\mathcal{D}}$, and we assume we cannot directly know their distribution, $p(X)$.

We then search for a function out of some known family that does "optimally" at answering a fixed question with a fixed framing.

- Classification: a yes-or-no or which-one-of-many question
- Regression: a continuous, how-much question

The major machine learning tasks

In machine learning, we mostly investigate one of several very broad tasks.

- Supervised learning: learn to answer a question by guessing, receiving the answer, and improving your guesses
- Reinforcement learning: learn to act on an environment by guessing actions, being rewarded or punished, and trying to earn rewards
- *Unsupervised learning*: learn to answer a question without being training on correct answers

Category Construction in Unsupervised Learning

What is concept learning in machine learning?

Concept learning is usually construed as the supervised or unsupervised learning of classifications. Categories are then considered to be the instances classified as belonging to the learned concept.

In engineering applications, supervised learning is most common.

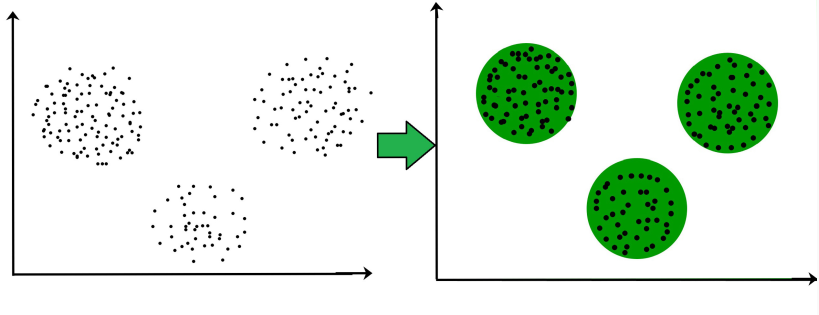
How is this unlike psychology and neuroscience?

- Supervised learning requires a human "teacher" to label data.
- This makes no pretense to capture ad-hoc concept construction,
- Nor embodied simulation,
- Nor can the concepts change with context, time, or bodily demand.
- But it has been commercially successful: ImageNet, AlphaGo Zero, face recognition, etc.

This limited achievement has been the real content of the recent boom in "artificial intelligence".

Unsupervised category construction?

The major unsupervised learning method for classification tasks is *clustering*. Given data in a feature space, you try to find “blobs” in the feature space. Any blob that does well by some metric can be labelled as a “cluster” and construed as a category.



What does this still lack?

Machine learning experts have begun to push for better unsupervised learning via *prediction* (eg. [10]). But (most) clustering methods do not predict over time, nor cluster as a function of context.

They cannot construct the kinds of dynamic, ad-hoc concepts we know the mind constructs.

Nor do they include the interoceptive features necessary for emotion concepts: bodily states, motivational valence, and a sense of agency.

Working back up: Concepts in Computational Cognitive Science

Taking the mind seriously

What if we started from what we know the brain and mind do, and tried to find which machine learning methods fit best?

Facts about the mind	Modeling methods
Sensory simulation	Generative models
Plausible inference	Probabilistic models (PGMs)
Goal-driven simulation	Bayesian conditioning in PGMs
Contextual simulation	Hierarchical PGMs
Prediction over time	Dynamical PGMs
Acting in the world	Causal PGMs
Universality	
Compositionality	*Probabilistic programs*

Table 1: Known facts about the mind as criteria for machine learning

Now we sound more like Noah Goodman[6] or Karl Friston[4, 5].

Ideas to drive PEN on concepts

- Hierarchical, dynamical probabilistic programs as simulators, with predictive coding as their neural implementation.
- Why have concepts? To learn, re-use, and compose simulators for novel situations, enabling us to survive and thrive.
- Compositional primitives make concept *construction* natural.

How are *emotion* concepts unique?

Emotion concepts predict *interoceptive* sensations: positive vs negative, not just predictable vs surprising. This defies the (purist) Free Energy Principle.

But they still have the features of other concepts:

- Valence and arousal commensurable within concept components,
- Affective concepts have compositionality,
- Affect can attach to anything causally connected to regulating the body.

This defies the usual model of valenced behavior in machine learning (reinforcement learning)

A pictorial example of affective concept composition

Conclusions: how do we bridge the gaps?

What mathematics can help with concepts?

- Probability: plausible, graded inferences,
- Category theory: compositional structures,
- Dynamical systems: prediction over time

Probabilistic programs can combine these, with practical applications.

Is that all? No.

We lack a compositional, dynamical formalism for affect and control. I would claim this is why "deep reinforcement learning doesn't work yet"[8].

Active inference model remains difficult and ad-hoc. Consider: Friston's models [5] vs. ForneyLab [3].

Active inference plays an important role in the Theory of Constructed Emotion[1].

Where to go from here?

We have a solid starting point to study how the mind constructs, composes, and re-uses ad-hoc concepts.

Computational grounding:

- Adaptor grammars[9]
- Probabilistic programs
- Dynamical Bayes nets

Psychological grounding:

- Embodied simulation
- Emotional granularity studies
- Studies of learning and development

Empirical ground:

- *Emotional granularity data*

Questions?

L. F. Barrett.
The theory of constructed emotion: an active inference account of interoception and categorization.
Social cognitive and affective neuroscience, 12(1):1–23, 2017.

L. W. Barsalou, W. K. Simmons, A. K. Barbey, and C. D. Wilson.
Grounding conceptual knowledge in modality-specific systems.
Trends in Cognitive Sciences, 7(2):84–91, 2003.

M. Cox, T. van de Laar, and B. de Vries.
A factor graph approach to automated design of Bayesian signal processing algorithms.
International Journal of Approximate Reasoning, 104:185–204, 2019.

- B. de Vries and K. J. Friston.
  A Factor Graph Description of Deep Temporal Active Inference.
  Frontiers in Computational Neuroscience, 11(October):1–16, 2017.
- K. J. Friston, R. Rosch, T. Parr, C. Price, and H. Bowman.
  Deep temporal models and active inference.
  Neuroscience and Biobehavioral Reviews, 77(April):388–402, 2017.
- N. D. Goodman, J. B. Tenenbaum, and T. Gerstenberg.
  Concepts in a Probabilistic Language of Thought.
  In *Concepts: New Directions*, number 010, pages 1–25. 2014.
- K. Grill-Spector and K. S. Weiner.
  The functional architecture of the ventral temporal cortex and its role in categorization.
  Nature Reviews Neuroscience, 15(8):536–548, 2014.

A. Irpan.
Deep reinforcement learning doesn't work yet.
`https://www.alexirpan.com/2018/02/14/rl-hard.html`, 2018.

M. Johnson, T. L. Griffiths, and S. Goldwater.
Adaptor grammars: A framework for specifying compositional nonparametric Bayesian models.
In *Neural Information Processing Systems (NIPS 2007)*, volume 19, page 641, 2007.

Y. Lecun.
Predictive Learning.
presented as keynote at Neural Information Processing Systems, 2016.